CSEF
Centre for Studies in Economics and Finance

WORKING PAPER NO. 201

Sorting the Good Guys from Bad: On the Optimal Audit Structure with Ex-Ante Information Acquisition

Anna Maria Cristina Menichini and Peter Simmons

July 2008

This version September 2013

University of Naples Federico II   University of Salerno   Bocconi University, Milan

CSEF - Centre for Studies in Economics and Finance
DEPARTMENT OF ECONOMICS – UNIVERSITY OF NAPLES
80126 NAPLES - ITALY
Tel. and fax +39 081 675372 – e-mail: csef@unisa.it

WORKING PAPER NO. 201

Sorting the Good Guys from Bad: On the Optimal Audit Structure with Ex-Ante Information Acquisition

Anna Maria Cristina Menichini[♥] **and Peter Simmons**[♦]

Abstract

In a costly state verification model under commitment the principal may acquire a costly public and imperfectly revealing signal before or after contracting. If the project remains profitable after all signal realisations, optimally the signal is collected, if at all, after contracting, and it may be acquired randomly or deterministically. Moreover, audit is deterministic and targeted on some signal-state combinations. The paper provides a detailed characterisation of the optimal contract and performs a comparative static analysis of signal acquisition strategy and payoffs with respect to enforcement costs and informativeness of the signal. We explore robustness of the results, including commitment issues.

Keywords: contracts, auditing, ex-ante information.

JEL codes: D82, D83, D86

Acknowledgements: We are grateful to two anonymous referees, Sandeep Baliga, Alberto Bennardo, Gabriella Chiesa, Marcello D'Amato, Paolo Garella, Tracy Lewis, Dilip Mookherjee, Marco Pagano, Nicola Persico, Salvatore Piccolo, Ailsa Roell, Maria Grazia Romano, Joel Sobel, Francesca Toscano, Zaifu Yang, the participants to the CSEF-IGIER Symposium on Economics and Institutions in Capri, the ASSET Conference in Padua, the RES Conference in Surrey, the 50th SIE Conference in Rome, the EARIE Conference in Toulouse, as well as the seminar participants at the universities of Bologna and Naples. The usual disclaimer applies.
A version of the paper has been circulated under the title "The Strategic and Social Power of Signal Acquisition".

[♥] Università di Salerno and CSEF. E-mail: amenichini@unisa.it

[♦] Department of Economics and Related Studies, University of York. E-mail: ps1@york.ac.uk

Table of contents

Introduction

In the costly state verification (CSV) models the reporting incentives of informed parties can often be controlled by costly audit. Within a commitment setting, in general audits should be stochastic (Border and Sobel, 1985; Mookherjee and Png, 1989). Deterministic audits may instead arise in a no commitment world to ensure that the contracted plan will be executed (Krasa and Villamil, 2000). However it begs the question of why audit alone is used to police cheating. Either informed or uninformed parties could seek alternatives. We analyse when and why it may be optimal to mix costly endogenous public signals with costly audit. In particular we allow uninformed parties to ex-ante spend resources to improve information regarding the likelihood of future events. Should they do this and if so when (before or after contracting)? Moreover we investigate, for each of these possible timings of information gathering, whether and how information alters the principal's ex-post audit strategy, providing a detailed characterisation of the optimal contract.

Many real world examples match our modelling scenario. In a corporate finance setting, before providing a capital injection, lenders may ask for a credit rating. The actual revenues that accrue after the loan are the managers' private information, but the lenders' decision to audit could be influenced by the rating company's ex ante assessment. In a tax audit scenario, tax authorities adopt a variety of methods to screen taxpayers, which may then affect the subsequent audit selection. In an insurance context, insurers collect information on policyholders characteristics, or history of past claims to assign them to credit risk classes. The insurer's decision to verify the occurrence or the magnitude of damages of policyholders filing for claims may be affected by the credit risk class they belong to.

We deal with the above questions by analysing the contract between an ex-post informed agent and a principal who can, at a cost, choose to become *ex-ante better* informed and/or *ex-post perfectly* informed. Specifically, we construct a CSV model in which, either before or after a contract is signed, but before the agents revenue state is realised, the principal can improve the information of both parties by acquiring a costly public signal correlated with the subsequent state.[1] This allows us to focus on the trade-off between ex-ante and ex-post information gathering

[1] We will henceforth refer to the ex-ante public information collected by the principal as a signal about the state, although it does not have the usual meaning of a privately informed person signalling (but see Section 8.2 for a discussion of this case).

as alternative policing methods available to the principal, the first corresponding to prevention and the second to punishment.

In this framework we find that, provided the information cost is not too high, it may be optimal to acquire information. However, the timing of information acquisition depends on the ex-ante profitability of the project. If this is positive after all signal realisations, information should be acquired only after the contract is signed, never before. This gives the advantage of using a broader ex-ante participation constraint for the principal and retaining flexibility in information acquisition, although it requires early commitment to investment. In these cases information acquisition occurs either randomly or deterministically, while the audit is deterministic and contingent - audit if the signal is good and do not audit if it is bad.

The frequency with which a signal is acquired is determined by the expected returns from such a strategy. If following signal collection returns are just sufficient to let the principal break even, then the signal is acquired for sure. If they exceed the minimal return demanded by the principal, then it is possible to save on signal acquisition cost by reducing its frequency and not auditing no signal states. If instead expected repayments following signal collection are insufficient for the principal to break even, then more resources can be raised by random signal acquisition accompanied by auditing for sure also no signal states as well.

If acquiring information is costless, the preferred audit strategy is no longer necessarily deterministic. There is an infinite number of optimal solutions, which involve equal audit cost and which range from deterministic audit and random signal acquisition to random audit and deterministic signal acquisition. In particular, a variation in the audit probability can be exactly compensated by a variation in the frequency of signal collection that keeps enforcement cost constant. This is not so with costly signals because they involve an ex-ante cost that is incurred irrespective of the true state and makes signal acquisition less attractive than auditing.

Lastly, if in some signal realisations the ex-ante profitability of the project is negative, it is no longer clear that committing to the investment before the signal dominates: getting the signal ex-ante allows the unprofitable projects to be ruled out, and, depending on the cost of the signal, may possibly make the overall profitability higher under ex-ante information acquisition. In such cases optimally auditing becomes stochastic again.

The widespread empirical practice of combining screening and costly audit indicates that the combination is a more efficient way of sorting than audit alone. We provide a theoretical understanding of when and why this is so. In tax audits, signals are increasingly used to help detect tax evasion and target audit policy as predicted by the model. The U.S. Internal Revenue Service (IRS), for example, uses its National Research Program (NRP) to compute expected tax liability for groups of taxpayers and uses this to identify individual tax returns that need a higher probability of audit.[2] Similar devices arise in tax systems using presumptive taxation methods, like France, Israel, Italy (see Thuronyi (1999) for a survey). In these systems, observable taxpayers characteristics correlated with real tax liability, like consumption patterns, are used to predict earnings and thus tax liability and to target audits on reports at odds with those inferred on the basis of the signal.

For expository reasons we use two states and two signal realisations and all the results are obtained in a commitment scenario. Extending the model to more than two states, we find that there is a tendency to pool higher income states, leading to a flat repayment in those states. This allows us to interpret our results in the light of observed features of financial contracts. In particular in our setting the principal/investor gets a debt contract, while the agent/entrepreneur gets a profit schedule resembling that of an equityholder.[3] As regards the commitment assumption, we find that, although deterministic audit overcomes the problems created by the unobservability of audit, a commitment problem reappears in implementing signal acquisition when it is random. We find that imposing sequential rationality leads to deterministic signal acquisition as well.

When the signal is acquired after the contract, the key results establishing deterministic auditing and random signal acquisition are stated in Propositions 1 and 3, followed by discussion of two special situations in Propositions 4 and 5, and by comparative static analysis of varying enforcement costs and informativeness of the signal on the signal acquisition strategy. A further key result establishes the optimal timing of signal acquisition (Proposition 8). The framework is extended to allow for more than two income states of the world (Propositions 9 to 11), no commitment (Proposition 12), private information and issues of distribution of the surplus. All proofs are in the

[2] The NRP has replaced the Taxpayer Compliance Measurement Program (TCMP), which consisted in more thorough audits of stratified random sample of taxpayers (for a description of both programmes, see Andreoni, Erand and Feinstein, 1998; Erard and Feinstein, 2010). Similar methods are used also in other countries.

[3] The more than two-state case is dealt with in Section 7, while the no commitment case in Section 8.1.

Appendix.

Related Literature. The paper is related to the literature on optimal audit strategy in CSV models. This can be split in two phases. The first only considers deterministic audit policies (Townsend, 1979; Gale and Hellwig, 1985) and finds that all states with revenues below some critical level should be audited with probability one, but that higher income states should never be audited and have a constant repayment. Reinganum and Wilde (1985) compare a common audit probability for all types with deterministic audit and find the latter is dominant. In the second phase, Border and Sobel (1987) and Mookherjee and Png (1989) point out that monitoring all defaulting states for sure is unduly expensive and that truthtelling may sometimes be achieved at a lower cost with random audit. Both studies use a setting with a finite number of states. Mookherjee and Png (1989), assuming the agent is risk-averse and his consumption is restricted to be positive in every state, find that audit should be stochastic and that the agent should be strictly rewarded for truthtelling. Border and Sobel (1987) use risk neutrality and find that the poorest individual gets no reward for truthtelling and in fact zero consumption so long as he is audited with positive probability. Moreover, they find one sufficient condition for the audit to be random: maximum possible reward for truthtelling (Proposition 3, p. 534). Chander and Wilde (1998) show that these results mainly extend to the case of a continuum of incomes.

All the above papers adopt a commitment principle: the principal announces an audit strategy sufficient to induce truthful reports by the agent and then sticks to it despite it being costly. At the interim stage there could nevertheless be a Pareto improving renegotiation or a non-cooperative solution altering the nature of the optimal audit strategy. While with a non-cooperative solution there is still scope for random auditing (Reinganum and Wilde, 1986; Khalil, 1997; Khalil and Parigi, 1998), a renegotiation which must be acceptable to all types (i.e., there is no interim possibility of revoking the contract) will instead make deterministic audit emerge as the optimal solution (Krasa and Villamil, 2000). In the present paper, the possibility of accessing costly ex-ante information is yet another reason for deterministic auditing in a commitment scenario and generates an efficiency gain relative to the standard ex-post only verification setup, complementing the efficiency gain approach in Krasa and Villamil (2000).

Besides the commitment assumption, another modelling difference with respect to Krasa and

Villamil (2000) lies in the focus on explicit information acquisition rather than implicit information revelation through actions (reports of the state in our framework). In Krasa and Villamil, by receiving a report about the state the principal updates her beliefs and on the basis of this chooses the enforcement strategy. In our setup, by receiving a signal which is positively correlated with the state the principal updates her beliefs about the true profitability of the project. However, since the signal is public and its value is exogenous to the contracting parties, there are no credibility/incentive problems about the information used for updating beliefs by the principal. In this sense the signal is a more effective discriminating device than an agent's report.

In conditioning audits on the realisation of an external signal correlated with true income, the paper is related to part of the tax literature, in particular Scotchmer (1987) and Macho Stadler and Perez Castrillo (2002). Within a setting in which information is freely available to the tax authority ex-ante, these papers determine the audit policy with linear taxes and fines and find that additional information can overcome the usual regressive bias of revenue maximising audit rules, where tax rates are highest on lower income groups or audit classes.

Last, in considering ex-ante information acquisition, the paper has some loose links with the literature on private or social value of precontractual costly information (Crémer and Khalil, 1992; Lewis and Sappington, 1997). In this literature the information can be privately gathered by the agent and perfectly reveals his type, thus altering the nature of the asymmetric information problem faced by the principal. In our paper the focus is rather different, as an imperfectly informative public signal is collected by the principal as part of her policing strategy.

1 The Model Assumptions

An agent has an investment project costing I for which he needs to raise funding from a risk neutral principal.[4] The project gives a random return f_s, $s \in \{H, L\}$, with $f_H > I > f_L > 0$, with probability π_H and π_L respectively. $I > f_L$ implies that to recover the investment cost the repayments to the principal (investors) must be non-decreasing in the state. While on their realisation the agent learns the revenues of the project for free, the principal can only discover them

[4]We model a financial contracting relationship, but the setup is general and can be adapted to apply to a variety of different contexts. Mookherjee and Png (1989) provide an interesting interpretation of a similar problem in a context of taxation.

with certainty by auditing at a cost $c_m \geq 0$. Audit is observable and the result of it is verifiable.

At any time before the report the principal can acquire a signal $\sigma \in \{G, B\}$ at a cost $c_a \geq 0$ which is positively correlated with the true state of nature. Signal acquisition is verifiable and contractible, moreover the realised signal value is public information. Denote with π_{ij} the joint probability that $s = i$ and $\sigma = j$, and $\pi_{i|j} = \Pr(s = i|\sigma = j)$ the conditional probability of state $s = i$ given that signal $\sigma = j$ is received. Let $\rho \equiv \pi_{HG}\pi_{LB} - \pi_{HB}\pi_{LG}$, with $\rho > 0$ defining the requirement of positive correlation ($r > 0$) between state and signal.[5] Under our assumptions, a good signal improves the chance that the state is actually high and vice versa.

We impose a condition which ensures that the ex-ante social benefits of the project cover the combined cost of any possible signal acquisition and monitoring policy:

Assumption 1

$$EP = \pi_H f_H + \pi_L f_L - I - \pi_L c_m - c_a > 0. \tag{1}$$

This assumption ensures that the project is ex-ante profitable even if enforcement involves both always getting the signal and monitoring every low state report.

We measure the timing of actions and events relative to the date at which a contract is agreed. We assume that the principal makes contract offers and that, among those, the agent chooses the one that maximises his expected utility. In line with Mookherjee and Png (1989), this is equivalent to assuming that the principal's problem is to find a scheme that maximises the utility of the agent subject to a minimum expected profit constraint and to the appropriate incentive constraints.[6] The implications of relaxing this assumption and letting the principal get all the surplus are discussed in Section 8.3.

The sequence of events is as follows: (1) The principal can acquire or not a signal. If acquired, the signal is publicly observed. (2) A financing contract is offered and, if accepted, the principal is committed to the investment. (3) The principal can acquire a signal (if she has not done so in

[5] The signal is fully informative if $\pi_{H|G} = \pi_{L|B} = 1$. In this case $r = 1$. Conversely, the signal is completely uninformative if $\pi_{H|G} = \pi_{H|B}$ and $\pi_{L|G} = \pi_{L|B}$ which implies $\pi_{HG} = \pi_{HB}$ and $\pi_{LG} = \pi_{LB}$ and then $r = \rho = 0$.

[6] This is then consistent either with the role of the principal as a utilitarian regulator who just wants to achieve an efficient monitoring system (e.g. a tax authority with a fixed revenue requirement who wishes to minimise the enforcement cost), or of a principal that maximises her own expected payoff but, due to competition between principals, is driven to offer contracts yielding her a zero expected payoff. Border and Sobel (1987) point out that for many objectives of the principal (the utilitarian case referred to above, minimising the expected audit cost or maximising the achievable expected return to the principal) an optimal scheme will be audit efficient.

stage 1). (4) The output is privately observed by the agent, who makes a report to the principal. (5) The principal can audit to discover the true value of output. (6) Payoffs are distributed.

A general game tree is sketched in Fig. 1.

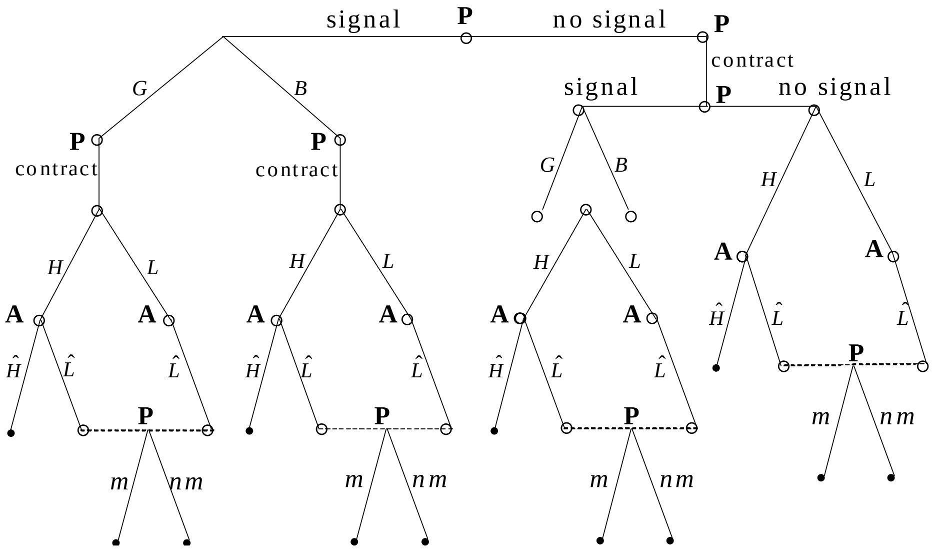


Fig. 1. The game tree

We refer to the two top branches of the figure as the left hand branch and the right hand branch. In each branch the variables written into the contract can depend on subsequent observable events and actions prior to the repayments actually being made. Hence $R_{\hat{s}s\sigma}$ is the repayment due following a signal value $\sigma \in \{G, B\}$, a report $\hat{s} \in \{\hat{H}, \hat{L}\}$, and an audit which reveals that the state is $s \in \{H, L\}$; $R_{\hat{s}\cdot\sigma}$ is the repayment with report $\hat{s}$ and signal σ, but with no audit; m_σ are the contracted probabilities of auditing a low report[7] following play of the signal strategy. In the right hand branch the contract also specifies a signal acquisition probability α, repayments $R_{\hat{s}sN}$, $R_{\hat{s}\cdot N}$ and monitoring probability m_N when play of signal strategy results in no signal acquisition. All repayments are non-negative and the agent has limited liability.

We first calculate the optimal contract and signal acquisition plan in the right hand branch of Fig. 1 and then repeat this for the left hand branch. Then, by comparing the expected values of these, we determine the optimal timing of information acquisition.

[7]In principle, these could vary with the reported state, but this does not occur since repayments are non-decreasing with the state. There is no incentive to cheat with low state incomes and hence no need to monitor high state repayment.

2 The contract problems conditional on timing of signal acquisition

Following the right hand branch of Fig. 1, a contract between principal and agent specifies repayments, audit probabilities and the probability with which information will be gathered.

The agent's payoff is

$$EU_A = \alpha \sum_\sigma \pi_\sigma U_\sigma + (1-\alpha) U_N \tag{2}$$

where

$$U_N \equiv \pi_H \left(f_H - R_{\hat{H}\cdot N}\right) + \pi_L \left[f_L - (1-m_N) R_{\hat{L}\cdot N} - m_N R_{\hat{L}LN}\right] \tag{3}$$

$$U_\sigma \equiv \pi_{H|\sigma} \left(f_H - R_{\hat{H}\cdot\sigma}\right) + \pi_{L|\sigma} \left[f_L - (1-m_\sigma) R_{\hat{L}\cdot\sigma} - m_\sigma R_{\hat{L}L\sigma}\right], \qquad \sigma \in \{G,B\}. \tag{4}$$

The principal's payoff is

$$E\Pi_P = \alpha \left\{\sum_\sigma \pi_\sigma PC_\sigma - c_a\right\} + (1-\alpha) PC_N - I \tag{5}$$

where

$$PC_N \equiv \pi_H R_{\hat{H}\cdot N} + \pi_L \left[(1-m_N) R_{\hat{L}\cdot N} + m_N \left(R_{\hat{L}LN} - c_m\right)\right] \tag{6}$$

$$PC_\sigma \equiv \pi_{H|\sigma} R_{\hat{H}\cdot\sigma} + \pi_{L|\sigma} \left[(1-m_\sigma) R_{\hat{L}\cdot\sigma} + m_\sigma \left(R_{\hat{L}L\sigma} - c_m\right)\right], \qquad \sigma \in \{G,B\}. \tag{7}$$

To induce truthful reporting, repayments following a truthful high state report, $R_{\hat{H}\cdot\sigma}$, $R_{\hat{H}\cdot N}$, must not exceed repayments following a false low state report which can be audited with probability m_σ, m_N :

$$TT_N \; : \; R_{\hat{H}\cdot N} \leq m_N R_{\hat{L}HN} + (1-m_N) R_{\hat{L}\cdot N} \tag{8}$$

$$TT_\sigma \; : \; R_{\hat{H}\cdot\sigma} \leq m_\sigma R_{\hat{L}H\sigma} + (1-m_\sigma) R_{\hat{L}\cdot\sigma}, \qquad \sigma \in \{G,B\}. \tag{9}$$

Last, from limited liability:

$$\begin{array}{c} f_H \geq R_{\hat{H}\cdot\sigma}, R_{\hat{L}H\sigma}, R_{\hat{H}\cdot N}, R_{\hat{L}HN} \\ f_L \geq R_{\hat{L}\cdot\sigma}, R_{\hat{L}L\sigma}, R_{\hat{L}\cdot N}, R_{\hat{L}LN} \end{array} \tag{10}$$

Signal acquired after the contract has been agreed

If the principal collects the signal after the contract, the contract P_{after} is a scheme $(\alpha, \{m\}, \{R\})$ that maximises the expected profits of the agent (2), subject to the principal's expected returns (5) meeting the investment outlay ($E\Pi_P \geq 0$), to the incentive constraints (8) and (9) and to the limited liability conditions (10).

Signal acquired before the contract has been agreed

If the principal collects the signal before the contract, the contract P_{b4} will be conditional on the received signal σ and will determine $(m_\sigma, \{R\})$ to maximise the agents' expected profits (4), subject to the principal expected return (7) meeting the investment cost ($PC_\sigma \geq I$), the truthtelling constraints (9) and the relevant limited liability conditions (10).

3 Generic properties of the optimal contract

Whatever the information gathering strategy, optimally each of the contracts $\mathcal{P}_{b4}, \mathcal{P}_{after}$ will display some common features (demonstrations are in the Appendix):

(i) the participation constraint must bind since otherwise it would be possible to reduce $R_{\hat{H}\cdot\sigma}, R_{\hat{H}\cdot N}$ without violating any of the constraints and make the agent better off;

(ii) the truthtelling constraints must all bind. After allowing for the participation constraint, the monitoring cost is a deadweight loss which ultimately subtracts from the expected gain to the agent. So, whatever the detailed structure of repayments, it is optimal to minimise the probability of monitoring. From this, it follows both that there must be maximum punishment for false audited reports, $R_{\hat{L}H\sigma} = f_H$, and that any relevant truthtelling constraints must bind.

(iii) in nonaudited low states the repayments $R_{\hat{L}\cdot\sigma}, R_{\hat{L}\cdot N}$ are set to give zero rent to the agent:[8] $R_{\hat{L}\cdot\sigma} = R_{\hat{L}\cdot N} = f_L$. This gives the agent the minimal incentive to cheat: if he cheats and is not actually audited, there is a chance he can profit at most by $f_H - f_L$.

(iv) in audited low states there is zero rent to the agent: $R_{\hat{L}L\sigma} = R_{\hat{L}LN} = f_L$. These repayments do not immediately impact on the incentive to cheat but, by setting them as high as possible, maximum revenue f_L for the principal from truthfully declared low states is realised in the participation constraint. This then allows a reduction in $R_{\hat{H}\cdot\sigma}, R_{\hat{H}\cdot N}$, and in turn a reduction in expected audit cost.[9]

[8] We take the lowest possible return to the agent to be 0. There may be institutional restrictions which limit punishment, e.g., bankruptcy legislation often requires creditors to leave a minimal standard of living to a defaulting individual. If it was set above this, largely we can just reinterpret f_L as being net of the minimum required return to the agent.

[9] With many states there is generally a reward for truthtelling in all but the lowest state (Border and Sobel, 1987; Mookherjee and Png, 1989). This is because paying a reward for truthful reporting in a state weakens the incentive

Using these properties, from the truth-telling constraints (8) and (9),

$$m_\sigma = \frac{R_{\hat{H}\cdot\sigma} - f_L}{f_H - f_L}, m_N = \frac{R_{\hat{H}\cdot} - f_L}{f_H - f_L}. \tag{11}$$

Thus, with commitment to audit, the contract will always ensure truthful reporting by the agent. Efficient policing is achieved by demanding maximum repayments in the low state.

The "reduced" contract problems

Having established the common properties of each contract, we use these to write the two contract problems in terms of just high state repayments.

If no signal is acquired before the contract, the contract problem becomes (P'_{after}):

$$\max_{\alpha, R_{\hat{H}\cdot\sigma}, R_{\hat{H}\cdot N}} \quad \alpha\left(\pi_H f_H - \textstyle\sum_\sigma \pi_{H\sigma} R_{\hat{H}\cdot\sigma}\right) + (1-\alpha)\,\pi_H\left(f_H - R_{\hat{H}\cdot N}\right)$$
$$\text{s.t. } \alpha\left\{\textstyle\sum_\sigma\left[\pi_{H\sigma} R_{\hat{H}\cdot\sigma} + \pi_{L\sigma}\left(f_L - m_\sigma c_m\right)\right] - c_a\right\} + (1-\alpha)\left\{\pi_H R_{\hat{H}\cdot N} + \pi_L\left(f_L - m_N c_m\right)\right\} = I \tag{12}$$

to the relevant limited liability constraints (10) and to $0 \le \alpha \le 1$, where m_σ, m_N are defined by (11).

If the signal is acquired before the contract then, when it has the value $\sigma \in \{G, B\}$, the contract problem $\mathcal{P}_{b4}$ becomes:

$$\max_{R_{\hat{H}\cdot\sigma}} \pi_{H|\sigma}\left(f_H - R_{\hat{H}\cdot\sigma}\right)$$
$$\text{s.t. } \pi_{H|\sigma} R_{\hat{H}\cdot\sigma} + \pi_{L|\sigma}(f_L - m_\sigma c_m) = I \tag{$\mathcal{P}'_{b4}$}$$
$$f_H \ge R_{\hat{H}\cdot\sigma} \ge f_L$$

We next determine the remaining properties of the optimal contract (high state repayments $R_{\hat{H}\cdot\sigma}, R_{\hat{H}\cdot N}$, and corresponding audit probabilities m_σ, m_N, and, in the right hand branch, signal acquisition strategy) with the different timings of signal acquisition, starting with the case in which a contract is agreed prior to the signal.

4 Signal strategy after the contract

If no information has been gathered before the contract, there are three possible forms of information gathering with associated high state repayments and audit after the contract:

for the agent to falsely declare the income of the adjacent lower state. However, for the lowest income state this incentive effect disappears and it is optimal, to maximise revenue collection, to give the lowest revenue agent zero consumption whether or not he is audited. The extension to more than two states is discussed in Section 7.

i. acquire the signal for sure;

ii. never acquire the signal;

iii. play a random strategy for acquiring the signal.

The choice between these depends on the relative costs of signal acquisition and monitoring, and on the informativeness of the signal. In order to derive the remaining properties of the contract, using (12) we rewrite the agent's expected profits as:

$$EU_{after} = \pi_H f_H + \pi_L f_L - I - \alpha \left(c_a + \textstyle\sum_\sigma \pi_{L\sigma} m_\sigma c_m\right) - (1-\alpha)\, \pi_L m_N c_m. \tag{13}$$

The contract problem can then equivalently be written as one of minimising the enforcement cost (expected cost of audit and signal acquisition) ($\mathcal{P}_{\mathcal{EC}}$):

$$\min EC \equiv \alpha \left(c_a + \textstyle\sum_\sigma \pi_{L\sigma} m_\sigma c_m\right) + (1-\alpha)\, \pi_L m_N c_m \tag{14}$$

subject to the same constraints as in problem $\mathcal{P}'_{after}$ ((10), (12) and $0 \leq \alpha \leq 1$).

4.1 Costly signal acquisition

When the signal is gathered either for sure or randomly at a cost c_a, a maximum spread between $R_{\hat{H}\cdot G}$ and $R_{\hat{H}\cdot B}$ is desirable. This is established in Proposition 1:

Proposition 1 *When $0 < \alpha \leq 1$, optimally $R_{\hat{H}\cdot G} = f_H$ and $R_{\hat{H}\cdot B} = f_L$. This implies that $m_G = 1$, $m_B = 0$.*

Proposition 1 has a pooling interpretation: after a bad signal, a flat repayment is collected independent of state ($R_{\hat{H}\cdot B} = R_{\hat{L}\cdot B} = R_{\hat{L}LB} = f_L$) and there is no audit ($m_B = 0$). After a good signal, repayments are state contingent and there is maximum spread between them ($R_{\hat{H}\cdot G} = f_H, R_{\hat{L}\cdot G} = R_{\hat{L}LG} = f_L$). This in turn implies that the principal always monitors after good signal realisations ($m_G = 1$), while she never does after bad signal realisations ($m_B = 0$). Intuitively, it is more efficient to require higher payments following a good signal since such payments can be supported with lower expected auditing costs. Indeed, since $\pi_{H|G} > \pi_{H|B}$ (and $\pi_{L|G} < \pi_{L|B}$), setting the spread in high state repayments $R_{\hat{H}\cdot G} - R_{\hat{H}\cdot B}$ as wide as possible both reduces audit costs, which are only paid in the low state, and makes the maximal contribution to repaying the investment cost in the participation constraint.

The agent obtains a rent following a bad signal when it turns out that revenue is actually high ($R_{\hat{H}\cdot B} = f_L < f_H$), but zero rent following a good signal ($R_{\hat{H}\cdot G} = f_H$). In some respects, leaving rent to the agent after a bad signal resembles debt forgiveness, although that usually arises in a no commitment scenario. For example, following the 2008 great recession, in the Eurozone both governments and banks have experienced debt forgiveness with various degrees of conditionality. Our analysis argues that it can be rational even in a commitment scenario to write debt forgiveness into the contract after a bad signal, leaving some rent to the debtor if the high state actually occurs.

Using Proposition 1, programme $P_{\mathcal{EC}}$ becomes $P'_{\mathcal{EC}}$:

$$\min \alpha \left(c_a + \pi_{LG} c_m\right) + (1-\alpha)\, \pi_L m_N c_m \tag{15}$$

$$\text{s.t. } \alpha EPC_\sigma + (1-\alpha)\left(PC_N - I\right) = 0 \tag{16}$$

where EPC_σ is the expected return to the principal upon signal collection, whether good or bad, while PC_N is the gross expected return when no signal is collected:

$$EPC_\sigma \equiv \pi_{HG} f_H + (1-\pi_{HG}) f_L - I - \pi_{LG} c_m - c_a \tag{17}$$

$$PC_N \equiv \pi_H R_{\hat{H}\cdot N} + \pi_L \left(f_L - m_N c_m\right) \tag{18}$$

This leaves the question of the optimal values of $R_{\hat{H}\cdot N}$ and α conditional on maximum spread values of $R_{\hat{H}\cdot G}$ and $R_{\hat{H}\cdot B}$. Remarkably, there is one simple condition, whose sign depends only on exogenous parameters, which captures the relative costs of signal acquisition and monitoring and the features of the joint probability distribution of signal and state:

$$NC_a \equiv \pi_H c_a - \rho c_m. \tag{19}$$

The expression NC_a is a measure of the net cost of signal acquisition and auditing. Intuitively, collecting the signal more frequently has a deadweight loss of $\pi_H c_a$ since with probability π_H there will be a truthful high state report anyway and no need for audit. But conversely more frequent signalling, allowing better targeting of audit, permits a saving on expected audit cost.[10]

[10] The cost saving to audit depends on how the informativeness of the signal leads to a differential between monitoring probabilities in good and bad signal states and can be written as $\rho c_m \left(m_G - m_B\right)$. This reduces to ρc_m as $m_G = 1, m_B = 0$. Consequently, the higher the auditing cost relative to the signal acquisition cost, the higher the gain from signal acquisition.

If $c_a = 0$, $NC_a \leq 0$ and we would expect the signal to be used. If $c_m = 0$, then $NC_a \geq 0$ and information acquisition becomes unattractive relative to monitoring. This is also the case if $\rho = 0$ so that the signal is completely uninformative. In the case of perfect correlation between signal and state, NC_a reduces to $\pi_H[c_a - \pi_L c_m]$.[11] Then, the sign of NC_a depends on the difference between the signal acquisition cost and the expected audit cost without signal acquisition. If positive, the deadweight loss of acquiring the signal exceeds the gain from saving audit costs and so it is not acquired.

According to the sign of NC_a, three possible cases arise. For sufficiently low auditing cost relative to signal acquisition cost, $NC_a > 0$: although signal acquisition buys precision in targeting repayments, this is so costly that the agent prefers to induce the principal not to gather the signal. The result gives stochastic monitoring.

Proposition 2 *When $NC_a > 0$, whatever the sign of EPC_σ, it is optimal to gather no information ($\alpha = 0$). The optimal contract has monitoring probability*

$$m_{N|\alpha=0} = \frac{I - f_L}{\pi_H (f_H - f_L) - \pi_L c_m} < 1, \tag{20}$$

high state repayment

$$R_{\hat{H}\cdot N}|_{\alpha=0} = \frac{(I - \pi_L f_L)(f_H - f_L) - f_L \pi_L c_m}{\pi_H (f_H - f_L) - \pi_L c_m} < f_H, \tag{21}$$

and cost function

$$EC_{\alpha=0} = \frac{(I - f_L)\pi_L c_m}{\pi_H (f_H - f_L) - \pi_L c_m}. \tag{22}$$

For a higher relative auditing cost, $NC_a < 0$: the signal is cheap enough and sufficiently informative to make it worth collecting ($\alpha \leq 1$). In this case, in determining the optimal combination of α, m_N, the expression NC_a (19) interacts with EPC_σ (17):

a. If $EPC_\sigma > 0$, the participation constraint (16) could be met entirely from with-signal states with $\alpha < 1$, not auditing without-signal states. Thus, the best mix of α and $R_{\hat{H}\cdot N}$ (m_N) involves reducing $R_{\hat{H}\cdot N}$ as far as possible to f_L, thus setting $m_N = 0$, and raising α as necessary to satisfy the participation constraint. This requires $\alpha = \underline{\alpha} < 1$ (defined in (23) below).

[11] In this case, $\pi_H = \pi_{HG}$, $\pi_L = 1 - \pi_{HG}$, $\rho = \pi_H \pi_L$.

b. If $EPC_\sigma = 0$, the return from the with-signal state fully meets the participation constraint. So the signal is acquired for sure ($\alpha = 1$), which means that $R_{\hat{H}\cdot N}$ is arbitrary.

c. If $EPC_\sigma < 0$, there is a shortage of revenue following signal acquisition. Since the participation constraint cannot be met solely from repayments arising from post-signal reports, more revenue must come from no signal states. The solution is to set $R_{\hat{H}\cdot N}$ at its highest possible rate (f_H), auditing all low reports in cases in which the signal is not acquired ($m_N = 1$). This generates the necessary revenue for the principal with relatively more infrequent signal acquisition α. The latter is obtained solving the participation constraint and is given by $\alpha = \bar{\alpha} < 1$ (defined in (26) below).

Proposition 3 summarises these findings:

Proposition 3 *When $NC_a < 0$ and the degree of correlation between signal and state is not too high, it is optimal to randomly or always gather information.*[12] *The actual values of $R_{\hat{H}\cdot N}$, m_N and α vary with the sign of EPC_σ. In particular:*

i. *(MS_a) If $EPC_\sigma > 0$, the unique optimum has $m_N = 0$, $R_{\hat{H}\cdot N} = f_L$,*

$$\alpha = \underline{\alpha} \equiv \frac{I - f_L}{\pi_{HG}\left(f_H - f_L\right) - \pi_{LG}c_m - c_a} < 1 \tag{23}$$

and cost function:

$$EC_{MS}^{a} = \frac{\left(I - f_L\right)\left(c_a + \pi_{LG}c_m\right)}{\pi_{HG}\left(f_H - f_L\right) - \pi_{LG}c_m - c_a}. \tag{24}$$

ii. *If $EPC_\sigma = 0$, it is optimal to always gather information ($\alpha = 1$). The cost function is*

$$EC_{\alpha=1} = c_a + \pi_{LG}c_m \tag{25}$$

iii. *(MS_b) If $EPC_\sigma < 0$, the unique optimum has $m_N = 1$, $R_{\hat{H}\cdot N} = f_H$,*

$$\alpha = \bar{\alpha} \equiv \frac{Ef - I - \pi_L c_m}{\pi_{HB}\left(f_H - f_L\right) - \pi_{LB}c_m + c_a} < 1 \tag{26}$$

and cost function

$$EC_{MS}^{b} = \frac{\left(\pi_H c_a - \rho c_m\right)\left(f_H - f_L\right) + \left(I - f_L\right)\left(\pi_{LB}c_m - c_a\right)}{\pi_{HB}\left(f_H - f_L\right) - \pi_{LB}c_m + c_a}. \tag{27}$$

[12] Namely $\rho \leq \frac{\pi_H \pi_L (I - f_L)}{\pi_{HG}(f_H - f_L)}$. For a higher degree of correlation, cases (*ii*) and (*iii*) might not arise. This is more thoroughly discussed in section 4.3.

Along with those in Proposition 1, these results show that, when the signal is worth acquiring it reverses the usual result about optimality of random audits and restores deterministic auditing even in situations in which parties are allowed to commit to stochastic auditing schemes. Depending on EPC_σ, repayments with no signal are either pooled with those after a bad signal (so $m_B = m_N = 0$, when revenues are relatively abundant) or pooled with the good signal repayments (so $m_G = m_N = 1$, when revenues are relatively scarce), while the signal strategy is set so as to meet the participation constraint. Intuitively, the reason for randomising on signal acquisition strategy rather audit is due to the fact that the signal has no impact on audit after a high state report but always has to be paid for, whereas the monitoring cost is zero after a high state report.[13]

When $NC_a = 0$, the principal is indifferent between not collecting the signal, getting $R_{\hat{H}\cdot N} = R_{\hat{H}\cdot N|\alpha=0}$ (21), or collecting it with positive probability and getting a repayment conditional on each signal-state combination. This is stated in Proposition 4:

Proposition 4 *If $NC_a = 0$, three cases can arise according to the sign of EPC_σ :*

i. *if $EPC_\sigma > 0$, all combinations of $m_N \in \left[0, m_{N|\alpha=0}\right]$ and $\alpha \in [\underline{\alpha}, 0]$ which satisfy the participation constraint (16) are possible;*

ii. *if $EPC_\sigma < 0$, all combinations of $m_N \in \left[m_{N|\alpha=0}, 1\right]$ and $\alpha \in [0, \bar{\alpha}]$ which satisfy the participation constraint (16) are possible;*

iii. *if $EPC_\sigma = 0$, there are two equally good solutions $(m_N, \alpha) = \{(0, m_N|_{\alpha=0}), (1, \mu)\}$ where $\mu \in [0, 1]$ is arbitrary.*

In all cases $m_{N|\alpha=0}, \underline{\alpha}, \bar{\alpha}$ are defined in (20), (23) and (26) respectively and the cost function is identical to that with no signal acquisition $EC_{\alpha=0}$ (22).

As Propositions 2 to 4 make clear, as NC_a varies from positive to negative values, the relative advantage of the signal increases and the policing policy varies continuously from exclusively auditing to combinations of audit and signal acquisition. The particular combinations are determined by how difficult it is to make the necessary repayments in different states. As NC_a

[13] When the signal is costless, this cost disadvantage of signal acquisition is lost and random audits can re-emerge (see Section 4.2). These arguments are more thoroughly developed in the Proofs of Propositions 3 and 5 in the Appendix.

passes through zero, the two policing methods have equal merit with comparable cost and so there is an element of substitutability between acquiring the signal for sure or never acquiring it.

4.2 Costless signal acquisition

Some special features arise when $c_a = 0$. In this case $NC_a < 0$, and so it is always optimal to acquire the signal. However, acquiring the signal and auditing tend to interact together in a one to one way on the enforcement cost. When $EPC_\sigma > 0$, these costs are $EC = \alpha\pi_{LG}m_G c_m$ so α and m_G can trade-off against each other. When $EPC_\sigma < 0$, $EC = [\alpha(\pi_{LG} + \pi_{LB}m_B) + (1-\alpha)\pi_L]\, c_m = [\pi_L - \alpha\pi_{LB}(1-m_B)]\, c_m$. Effectively, α and $1 - m_B$ appear as perfect substitutes in EC. In both cases this generates multiple optimal enforcement policies. Proposition 5 summarises these findings:

Proposition 5 *If $c_a = 0$, and still the degree of correlation between signal and state is not too high, the repayment structure and the signal dependent monitoring probabilities depend on the sign of EPC_σ :*

i. when $EPC_\sigma > 0$, $m_N = m_B = 0$, $R_{\hat{H}\cdot N} = R_{\hat{H}\cdot B} = f_L$. There are infinitely many optimal possibilities with $\underline{\alpha} \leq \alpha \leq 1$ and $m_G \leq 1$:

$$\begin{aligned} R_{\hat{H}\cdot G} &= f_L + \frac{(I-f_L)(f_H-f_L)}{\alpha[\pi_{HG}(f_H-f_L) - \pi_{LG}c_m]} \leq f_H \\ m_G &= \frac{I-f_L}{\alpha[\pi_{HG}(f_H-f_L) - \pi_{LG}c_m]} \leq 1 \end{aligned}$$

The cost function coincides with EC^a_{MS} in (24) when $c_a = 0$.

ii. when $EPC_\sigma = 0$, $m_G = 1$, $R_{\hat{H}\cdot G} = f_H$, $m_B = 0$, $R_{\hat{H}\cdot B} = f_L$. It is optimal to always gather information ($\alpha = 1$) and the cost function coincides with $EC_{\alpha=1}$ in (25) when $c_a = 0$.

iii. when $EPC_\sigma < 0$, $m_G = m_N = 1$ and $R_{\hat{H}\cdot G} = R_{\hat{H}\cdot N} = f_H$. There are infinitely many optimal possibilities with $\bar{\alpha} \leq \alpha \leq 1$, and $0 \leq m_B < 1$:

$$\begin{aligned} R_{\hat{H}\cdot B} &= f_H - \frac{(f_H-f_L)(\pi_H f_H + \pi_L(f_L-c_m) - I)}{\alpha\,[\pi_{HB}(f_H-f_L) - \pi_{LB}c_m]} < f_H \\ m_B &= 1 - \frac{\pi_H f_H + \pi_L(f_L-c_m) - I}{\alpha\,[\pi_{HB}(f_H-f_L) - \pi_{LB}c_m]} \geq 0 \end{aligned}$$

The cost function coincides with EC^b_{MS} (27) when $c_a = 0$.

These results are in striking contrast with those obtained under costly signal acquisition, where audits are deterministic and randomisation occurs on the signal acquisition side. Under costless signal acquisition there is some indifference between randomising over the signal acquisition strategy and some of the audit strategies. As highlighted in Section 4.1, the difference between the two cases is to be found in the presence of the signal acquisition cost c_a. A costly signal has a disadvantage relative to audit, because it is paid even when it is not needed, but when $c_a = 0$ audit and signal are equally costly means of enforcement.

The solutions just seen under costly and costless signal collection are sketched in Fig. 2, which divides the $c_a - c_m$ space into areas according to the sign of NC_a and EPC_σ.

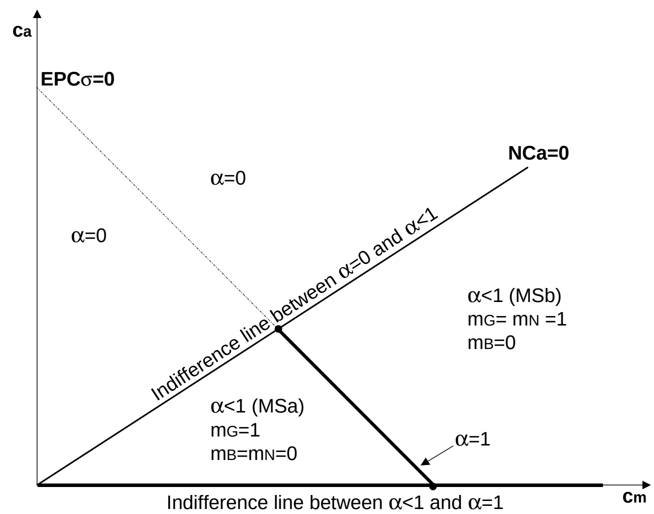


Fig. 2. Signal and audit strategy in the space of audit and signal acquisition cost.

To the left of the $NC_a = 0$ line, the signal is never acquired, regardless of the sign of EPC_σ. To the right of it, the signal is acquired either randomly or for sure depending on the sign of EPC_σ. Last, the $NC_a = 0$ and the $c_a = 0$ lines denote indifference regions: along the $NC_a = 0$ line the principal is indifferent between acquiring the signal or not, while along the $c_a = 0$ line she is indifferent to acquiring it for sure or randomly.

This shows that the "standard" stochastic auditing result need not be optimal when it is possible for either party to unilaterally improve the information structure of both ex-ante.[14]

[14] The propositions also hold if it is the agent who can collect the signal.

4.3 Informativeness of the signal

The above results have been derived assuming that the signal is not too informative. However, a high degree of correlation between signal and state might rule out some of the solutions listed in Propositions 2 to 5. To see why, note that the overall project profitability expression EP (1) can be decomposed into the sum of two factors: EPC_σ (17), which is a measure of the profitability of the project in with-signal states, and $EP_B = \pi_{HB}(f_H - f_L) - \pi_{LB}c_m$, which is a measure of the gain to the principal from monitoring low reports.

Consider now the three relations $EP = EPC_\sigma = EP_B = 0$ in $c_a - c_m$ space. These are linear and their slopes are such that the line $EP_B = 0$ is vertical, while $EP = 0$ and $EPC_\sigma = 0$ are each downward sloping, with EP steeper than EPC_σ. Since $EP = EPC_\sigma + EP_B$, any two of these three lines intersect at

$$c_m^* = \frac{\pi_{HB}}{\pi_{LB}}(f_H - f_L)\,;\; c_a^* = \frac{\rho}{\pi_{LB}}(f_H - f_L) + f_L - I \tag{28}$$

Since $EP > 0$ from Assumption 1, our feasible set is confined to the area below the $EP = 0$ line (the red solid downward sloping line in Figures 3a to 3d). The signal is of value only if $NC_a \leq 0$ so another intersection of interest is that between EPC_σ and NC_a which occurs at

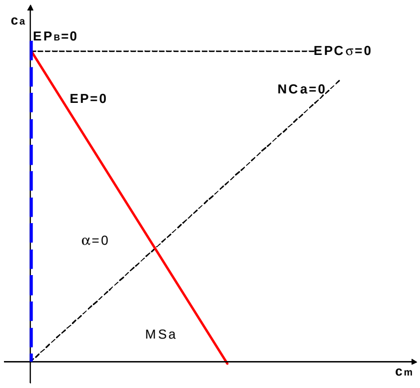
$$\widehat{c}_m = \frac{\pi_H(\pi_{HG}f_H + (1-\pi_{HG})f_L - I)}{\pi_{HG}\pi_L};\; \widehat{c}_a = \frac{\rho(\pi_{HG}f_H + (1-\pi_{HG})f_L - I)}{\pi_{HG}\pi_L} \tag{29}$$

The signal acquisition strategies MS_b and $\alpha = 1$ can only arise if there are points c_a, c_m with $EP < 0, EPC_\sigma \leq 0$ and $NC_a \leq 0$. If $\widehat{c}_a < c_a^*$ no such points exist although the solution MS_a is still possible. But if $\widehat{c}_a > c_a^*$ such points do exist and there are values of the enforcement costs supporting each of the signal strategies in Proposition 3. The difference $c_a^* - \widehat{c}_a$ (which has the opposite sign of $c_m^* - \widehat{c}_m$) is equal to $\rho\pi_{HG}(f_H - f_L) - \pi_H\pi_{LB}(I - f_L)$. It is increasing in ρ and zero when $\rho = \rho^* \equiv \frac{\pi_H\pi_{LB}(I-f_L)}{\pi_{HG}(f_H-f_L)}$. So $\rho < \rho^*$ allows all signal strategies to occur whereas with $\rho > \rho^*$ the only possible acquisition strategy given $NC_a \leq 0$ is MS_a.

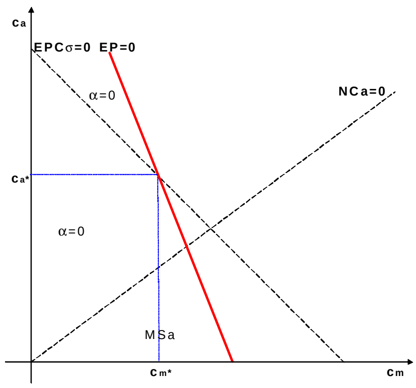
Proposition 6 *Define* $\rho^* \equiv \frac{\pi_H\pi_{LB}(I-f_L)}{\pi_{HG}(f_H-f_L)}$*. If it is worth collecting the signal, when* $\rho < \rho^*$*, the results of Proposition 3 hold, but if* $\rho \geq \rho^*$*, only case* MS_a *is optimal.*

It is easy to see this graphically. Depending on the degree of correlation between signal and state four possible scenarios can arise. When the correlation is 1, $\pi_{HG} = \pi_H$, $\pi_{LB} = \pi_L$, $\pi_{HB} = \pi_{LG} = 0$.

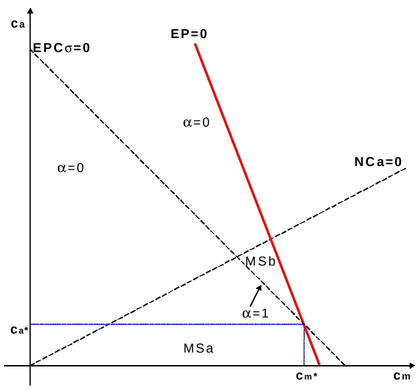
The EP_B line coincides with the vertical axis and the EPC_σ line is horizontal and has the same intercept as $EP = 0$. Since the feasible set is confined to the area below the $EP = 0$ line, the possible solutions are described by $\alpha = 0$ in Prop. 2 and by case MS_a in Prop. 3 (fig. 3a). As the correlation falls, EP_B shifts rightward. For high but not perfect correlation, the solution form is still described by $\alpha = 0$ or by case MS_a (fig. 3b). As the correlation falls further, there is a region of correlation for which all of the cases described in Propositions 2 to 5 enter as a possible solution ($\alpha = 0$, MS_a, MS_b, $\alpha = 1$) (fig. 3c). When the correlation is zero (fig. 3d), the signal has no information value, the NC_a line coincides with the horizontal axis and the unique optimal solution has $\alpha = 0$ (Proposition 2).



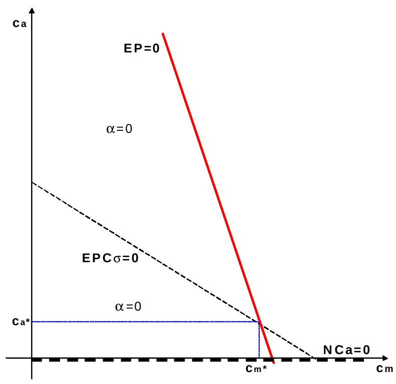
3a. Perfectly informative signal



3b. Highly informative signal



3c. Mildly informative signal



3d. Uninformative signal

4.4 Signal strategy as a function of acquisition and audit costs

The results in Propositions 2 to 5 allow us to analyse how the signal acquisition strategy varies with c_a and c_m. The striking result here is that the frequency of signal acquisition is non-monotone in audit and signal collection costs. When $NC_a < 0$ and $EPC_\sigma \geq 0$, the probability of signal acquisition increases at an increasing rate with audit and signal acquisition costs. On the other hand, when $EPC_\sigma < 0$, the optimal α starts decreasing in both audit (at a decreasing rate) and signal acquisition cost (at an increasing rate). The $\alpha(c_m, c_a)$ function is continuous (except when $NC_a = 0$) and has a kink at its maximum, which occurs where $EPC_\sigma = 0$. At first sight it is surprising that α can increase with c_a. Starting from a point at which MS_a is optimal (and so $m_N = 0$) (case i, Prop. 3), there is a large saving from not auditing no signal states. But as c_a rises, to satisfy the participation constraint, some extra revenue must be generated and this is done by increasing α. If initially c_m is high enough,[15] as c_a rises α can increase all the way to $\alpha = 1$. For lower initial values of c_m, α can only rise so far. Eventually, for any initial value of c_m, as c_a gets high enough, α falls discontinuously to zero. This pattern is depicted in figure 4.

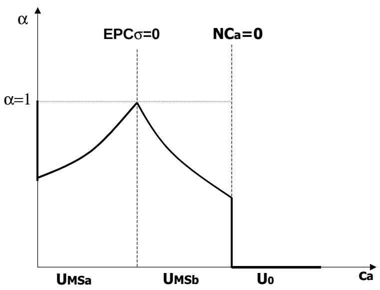


Fig. 4. $\alpha(c_a)$ for $c_m > 0$.

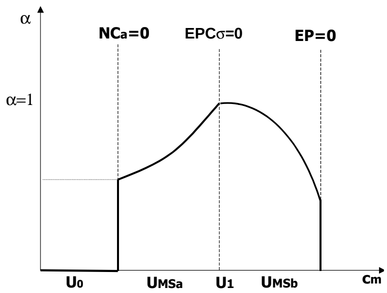


Fig. 5. $\alpha(c_m)$ for $c_a > 0$

A similar pattern arises if we increase c_m, keeping c_a fixed, as depicted in figure 5. In this case, if c_a is low enough,[16] α can increase all the way to $\alpha = 1$, and then decreases until c_m is so high to violate Assumption 1.

Obviously, utility falls as signal acquisition and audit cost rise. It falls continuously, with kinks when the sign of NC_a and EPC_σ changes, as Figures 8 and 9 in Appendix show.

5 Signal gathered before the contract

We now turn to the case in which a signal is collected before any contractual commitment. The peculiarity of this setting is that collecting the signal ex-ante may impact on the principal's decision to make a contract offer. In particular, since the signal is realised before any contractual

[15] I.e., for c_m no less than the value at which EPC_σ and NC_a intersect: $c_m \geq \widehat{c}_m$.

[16] I.e., for $c_a \leq \widehat{c}_a$, the value at which EPC_σ and NC_a intersect.

commitment, if it gives sufficiently bad news about the project, the investor can decide to not undertake it. This depends on the expected profits after a bad signal:

$$\Pi_B \equiv \pi_{H|B} f_H + (1 - \pi_{H|B}) f_L) - I - \pi_{L|B} c_m. \tag{30}$$

If $\Pi_B \leq 0$, the principal will make no contract offer upon receiving a bad signal.

Solution to Program P_{b4} is summarised by the following proposition:

Proposition 7 *When the contract is agreed after the signal has been collected,*

(i) *if* $\Pi_B > 0$, *the project is executed after any signal and the contract stipulates:*

$$m_{\sigma|b4} = \frac{I - f_L}{\pi_{H|\sigma}(f_H - f_L) - \pi_{L|\sigma} c_m}, \tag{31}$$

with $m_{G|b4} < m_{B|b4}$; *high state repayments*

$$R_{\hat{H} \cdot \sigma} = \frac{(f_H - f_L) I - \pi_{L|\sigma} f_L (f_H - f_L + c_m)}{\pi_{H|\sigma}(f_H - f_L) - \pi_{L|\sigma} c_m}; \tag{32}$$

and the agent's ex ante utility

$$EU_{b4} = \textstyle\sum_\sigma \pi_\sigma \left\{ \pi_{H|\sigma} f_H + \pi_{L|\sigma} f_L - I - \dfrac{\pi_{L|\sigma} c_m (I - f_L)}{\pi_{H|\sigma}(f_H - f_L) - \pi_{L|\sigma} c_m} - c_a \right\}; \tag{33}$$

(ii) *if* $\Pi_B \leq 0$, *the project is abandoned after a bad signal but executed after a good signal and the contract stipulates audit probability and high state repayment conditional on a good signal defined in 31 and 32 respectively, and the agent's ex ante utility is*

$$EU_{G|b4} = \pi_G \left\{ \pi_{H|G} f_H + \pi_{L|G} f_L - I - \frac{\pi_{L|G} c_m (I - f_L)}{\pi_{H|G}(f_H - f_L) - \pi_{L|G} c_m} \right\} - c_a. \tag{34}$$

The above proposition shows that if the signal is gathered before investment is committed, the contract has the same qualitative properties as the CSV commitment contract with no information gathering described in Proposition 2, in particular stochastic monitoring conditional on the signal re-emerges.

One interesting point is that, provided that the expected return conditional on a bad signal realisation is positive, audits after a bad signal occur with a higher frequency than audits after a good signal ($m_{B|b4} > m_{G|b4}$) in contrast to the results we have in the right hand branch where $m_G = 1 > m_B = 0$. This is because here the principal's participation constraint is in ex-post terms

- it must be satisfied state-by state - and, to ensure compliance, auditing has to be carried out in each state-signal realisation. Since receiving a good signal makes the high state more likely, expected revenues from nonaudited high states increase. This slackens the participation constraint, reduces the repayment $R_{\hat{H}\cdot G}$ to the principal and lessens the need for audit of low reported states after a good signal. The opposite is true if a bad signal is received.

6 Globally optimal signal strategy

To determine when to collect the signal, the principal compares the expected return when the signal is collected before any contract offer, (33) or (34), according to the sign of Π_B (30), with that obtained when the signal is collected after a contract offer is made (13).

Suppose first that $\Pi_B > 0$, i.e., it is profitable to offer a contract even after a bad signal. Coupled with $f_L < I$, this assumption sets an upper bound on the maximum informativeness of the signal, precluding the case in which the signal is perfectly informative ($\pi_{H|B} = \pi_{L|G} = 0$) and crowds out audit. This leads to the following proposition:

Proposition 8 *Under the assumption that* $\Pi_B > 0$, *when* $NC_a < 0$ *it is never optimal to gather the signal before a contract has been agreed.*

The principal will always strictly prefer to agree to a contract prior to any signal acquisition. So the right hand branch of the game tree strictly dominates the left hand branch. This is because the right hand branch saves signal costs and allows the participation constraint (16) to be set in ex-ante terms. The possibility of cross-subsidisation between these different signal states means that if the signal is collected for sure, the right hand branch must always at least weakly dominate the left hand branch. That is, within the right hand branch we can replicate the left hand branch for any particular signal value by setting $\alpha = 1$. If $EPC_\sigma = 0$, $\alpha = 1$ is actually optimal in the right hand branch and there is cross-subsidisation from the profitable good signal outcome to the bad signal outcome. Then the right hand branch strictly dominates the left hand one. In other cases, when optimally $\alpha < 1$ in the right hand branch, the optimal right hand branch solution again must strictly dominate the solution restricted to $\alpha = 1$ and so must dominate the left hand branch.

Assuming $\Pi_B > 0$ limits the range of correlation between signal and state and allows us to derive a clear superiority of setting the contract after getting the signal rather than before. However it also rules out one useful role for the signal in detecting projects which are ex-ante unprofitable and from which the investor would like to abstain. The literature which does consider signalling highlights this role for ex-ante information acquisition (Crémer, 1992). So suppose that conditional on a bad signal realisation the ex-ante expected return from the project is negative $\Pi_B < 0$. If the signal is collected prior to commitment to the investment loan, then with a bad signal realisation the project will not be undertaken (with a good signal of course the project is revealed to be more profitable than at first sight and so will be undertaken). On the other hand, if the signal is collected after commitment to the investment, then this opportunity of not undertaking a project revealed to have a negative ex-ante return is sacrificed.

However, it is not always the case that the left hand branch dominates. The intuition for this is that the expected gain from the project for a particular time sequence of contract writing and signal acquisition is composed of three distinct elements: the expected gross revenue that the project will yield allowing for optimal decision making; the expected monitoring cost; the expected signalling cost. The expected signal cost is generally lower if the signal is collected only after the investment has been committed since both cases MS_a and MS_b in Proposition 3 have $\alpha < 1$. However, the expected audit cost in the different timings is ambiguous and depends on whether the optimal audit strategy when investment is committed first corresponds to MS_a, MS_b or $\alpha = 1$. The expected audit cost in the different timings is ambiguous and depends on whether the optimal audit strategy when investment is committed first corresponds to MS_a, MS_b or $\alpha = 1$. The expected revenues of the project in the different timings are also ambiguous, their difference being $\pi_{H|B} f_H + (1 - \pi_{H|B}) f_L - I$. Which of these effects dominate then depends on the precise values of c_a, c_m and the probability distribution of project returns.[17]

For a particular probability distribution of profits ($f_H = 10, f_L = 1, \pi_{HB} = 0.1, \pi_{LG} = 0.1, I = 4$ which implies that $\rho = 0.11$), Fig. 6 ($\pi_{HG} = .6$) and 7 ($\pi_{HG} = .4$) plot combinations of c_m, c_a giving indifference between the left and right hand branches for $EPC_\sigma > 0$ and $EPC_\sigma < 0$ respectively.

[17] The ambiguity remains even in the extreme case in which the signal is perfectly informative.

Inside the area delimited by the blue curve, the right hand branch still dominates.

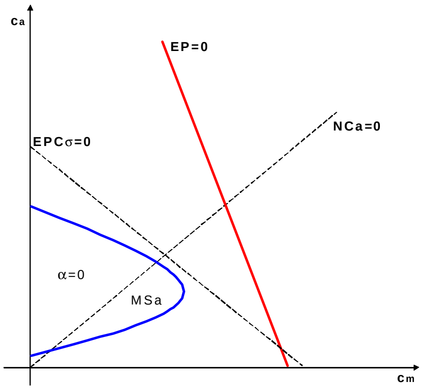


Fig. 6. Region of optimality of right hand branch when $EPC_\sigma > 0$.

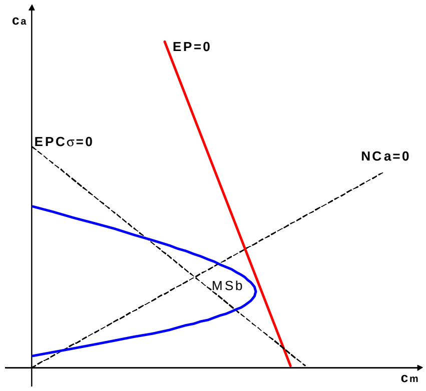


Fig. 7. Region of optimality of right hand branch when $EPC_\sigma < 0$.

Another interesting comparison can be made for the case in which $NC_a > 0$ in the right hand branch. In this case the signal is never collected after investment is committed ($\alpha = 0$) and the optimal audit strategy is random. Again, despite the fact that ex-ante signal acquisition makes it possible to rule out unprofitable projects, the right hand branch may still dominate the left hand branch if signal saving cost is high.

7 More than two income levels

We now show that the results of the two-state world apply more generally with many income states for suitable probability distributions of revenues. Suppose revenues f can take three possible realisations, high, medium and low (H, M, L), with $f_H > f_M > I > f_L$, and suppose that prior to the realisation of the state the principal can acquire a signal which can take two realisations, good and bad (G, B).

The probability distribution of states and signals is given by

	G	B	Σ_s
H	π_{HG}	π_{HB}	π_H
M	π_{MG}	π_{MB}	π_M
L	π_{LG}	π_{LB}	π_L
Σ_σ	π_G	π_B	1

We assume that the signal and the state satisfy the bivariate monotone likelihood property ($BMLP$) (Chew, 1973), which reflects the idea of positive correlation between signal and state. In particular, it implies that each 2x2 minor of the 3x2 probability matrix is positive, i.e.

$$\begin{aligned}\pi_{HG}\pi_{MB} &> \pi_{HB}\pi_{MG},\\ \pi_{HG}\pi_{LB} &> \pi_{HB}\pi_{LG},\\ \pi_{MG}\pi_{LB} &> \pi_{MB}\pi_{LG}.\end{aligned}$$

$BMLP$ implies:

$$\rho_3 \equiv (\pi_{HG}+\pi_{MG})\pi_{LB} - (\pi_{HB}+\pi_{MB})\pi_{LG} > 0 \tag{35}$$

and has the interpretation of the chance of one of the top two states being higher after a good than a bad signal.

Assumption 1 is now replaced by a stronger condition:[18]

$$EP_3 \equiv (\pi_H+\pi_M)\, f_M + \pi_L f_L - I - \pi_L c_m - c_a > 0. \tag{36}$$

Essentially a distribution in which the top two states have the same revenue covers the investment and agency cost. EP_3 can be decomposed into the sum of $EPC_{\sigma_3} \equiv (\pi_{HG}+\pi_{MG})\, f_M + (1-\pi_{HG}-\pi_{MG})\, f_L - I - \pi_{LG}c_m - \alpha c_a$ and $EP_{B_3} \equiv (\pi_{HB}+\pi_{MB})\,(f_M - f_L) - \pi_{LB}c_m$.

Let $R_{\hat{s}s\sigma}$ be the repayment due following a signal value $\sigma \in \{G, B\}$, a report $\hat{s} \in \{\hat{H}, \hat{M}, \hat{L}\}$, and an audit which reveals that the state is $s \in \{H, M, L\}$, $R_{\hat{s}s}$ the analogous repayment with no signal acquisition, $R_{\hat{s}\cdot\sigma}$ the repayment with report $\hat{s}$ and signal σ, but with no audit, and $R_{\hat{s}\cdot}$ the analogous repayment with no signal acquisition. The facts that an optimal scheme must involve maximum punishment and also a zero rent to the agent in the lowest state are independent of the numbers of state and signal realisations (Border and Sobel, 1987), so $R_{\hat{L}L} = R_{\hat{L}\cdot} = R_{\hat{L}L\sigma} = R_{\hat{L}\cdot\sigma} = f_L$, and $R_{\hat{M}H} = R_{\hat{M}H\sigma} = R_{\hat{L}H} = R_{\hat{L}H\sigma} = f_H; R_{\hat{L}M} = R_{\hat{L}M\sigma} = f_M$.

Letting $m_{s\sigma}, m_s$ be the probability of auditing a report of states $s \in \{L, M\}$ and signal $\sigma \in \{G, B\}$, the incentive compatibility conditions ensuring truthful reporting when the state is high or medium become:

- for the high type to truthfully report high and not medium ($IC\,(HM)$):

$$\begin{aligned}\left(f_H - R_{\hat{M}\cdot}\right)(1-m_M) &\leq f_H - R_{\hat{H}\cdot} \\ \left(f_H - R_{\hat{M}\cdot\sigma}\right)(1-m_{M\sigma}) &\leq f_H - R_{\hat{H}\cdot\sigma}\end{aligned} \tag{37}$$

[18] It is stronger because $(\pi_H+\pi_M)\, f_M + \pi_L f_L - \pi_H f_H - \pi_M f_M - \pi_L f_L = \pi_H\,(f_M - f_H) < 0$.

- for the high type to truthfully report high and not low ($IC\,(HL)$):

$$\begin{aligned}(f_H - f_L)(1 - m_L) &\leq f_H - R_{\hat{H}\cdot} \qquad (38)\\ (f_H - f_L)(1 - m_{L\sigma}) &\leq f_H - R_{\hat{H}\cdot\sigma}\end{aligned}$$

- for the intermediate type to truthfully report medium and not low ($IC\,(ML)$):

$$\begin{aligned}(f_M - f_L)(1 - m_L) &\leq f_M - \left[(1 - m_M) R_{\hat{M}\cdot} + m_M R_{\hat{M}M}\right] \qquad (39)\\ (f_M - f_L)(1 - m_{L\sigma}) &\leq f_M - \left[(1 - m_{M\sigma}) R_{\hat{M}\cdot\sigma} + m_{MG} R_{\hat{M}M\sigma}\right]\end{aligned}$$

The contract problem is to minimise the expected signal acquisition cost and audit cost after a low or intermediate report

$$\min_{\alpha,\{m\},\{R\}} \alpha\Sigma_\sigma\left[(\pi_{M\sigma}m_{M\sigma} + \pi_{L\sigma}m_{L\sigma})\,c_m + c_a\right] + (1-\alpha)(\pi_M m_M + \pi_L m_L)\,c_m \qquad (40)$$

subject to the incentive constraints above and the participation constraint being non-negative:

$$\begin{aligned}&\alpha\Sigma_\sigma\left\{\pi_{H\sigma}R_{\hat{H}\cdot\sigma} + \pi_{M\sigma}\left[(1-m_{M\sigma})R_{\hat{M}\cdot\sigma} + m_{M\sigma}R_{\hat{M}M\sigma}\right] - (\pi_{M\sigma}m_{M\sigma} + \pi_{L\sigma}m_{L\sigma})\,c_m - c_a\right\} \qquad (41)\\ &+ (1-\alpha)\left\{\pi_H R_{\hat{H}\cdot} + \pi_M\left[(1-m_M)R_{\hat{M}\cdot} + m_M R_{\hat{M}M}\right] - (\pi_M m_M + \pi_L m_L)\,c_m\right\} + \pi_L f_L - I\end{aligned}$$

Minimum monitoring costs are achieved with binding incentive constraints (37) and (39). Together then these imply that (38) is slack. In addition we can show that for each signal situation, $m_{M\sigma} = m_M = 0$ so that optimally reports of the top two states are not audited at all, and they have a common repayment which is at most equal to f_M. The rationale is that if $m_M, m_{M\sigma} > 0$, it is possible to raise $R_{\hat{M}\cdot}, R_{\hat{M}\cdot\sigma}$ up to f_M, and if necessary reduce $R_{\hat{H}\cdot}, R_{\hat{H}\cdot\sigma}$ down to f_M whilst reducing $m_M, m_{M\sigma}$ to zero, thus reducing audit costs. The fact that the top two states are pooled is a key result in characterising the optimal contract.

Proposition 9 *If EP_3 holds, optimally (37) and (39) bind, $m_M, m_{M\sigma} = 0$, $R_{\hat{M}\cdot} = R_{\hat{H}\cdot}$, $R_{\hat{H}\cdot\sigma} = R_{\hat{M}\cdot\sigma}$. Thus, the optimal audit scheme involves pooling the top two states.*

The argument for pooling the top two states is applicable to any problem with more than two states. Hence with n states it could be applied successively, involving pooling over the top $n-1$ states so long as:

$$EP_n = \Sigma_{i=2}^n \pi_i f_2 + \pi_1 f_1 - I - \pi_1 c_m - c_a > 0$$

i.e., in general there is sufficient revenue in the lowest non-audited state.

Pooling states H and M, the three-state problem reduces to one in two new states. The issues then mirror the two state arguments with suitable modifications. The questions become finding the best split of upper state repayments, low report audit probability and signal acquisition probability. Imposing $m_M = m_{M\sigma} = 0$ and $R_{\hat{M}\cdot} = R_{\hat{H}\cdot}, R_{\hat{M}\cdot\sigma} = R_{\hat{H}\cdot\sigma}$, the problem becomes $(\mathcal{P}'_{\mathcal{EC}_3})$:

$$\min \alpha \left(\Sigma_\sigma \pi_{L\sigma} m_{L\sigma} c_m + c_a\right) + (1-\alpha)\,\pi_L m_L c_m \tag{42}$$

$$\begin{aligned} &\text{s.t. } \alpha\Sigma_\sigma \left\{(\pi_{H\sigma} + \pi_{M\sigma})\, R_{\hat{M}\cdot\sigma} - \pi_{L\sigma} m_{L\sigma} c_m - c_a\right\} + \\ &+ (1-\alpha)\left\{(\pi_H + \pi_M)\, R_{\hat{M}\cdot} - \pi_L m_L c_m\right\} + \pi_L f_L = I \end{aligned} \tag{43}$$

and the single binding incentive constraint (39)

$$\begin{aligned} R_{\hat{M}\cdot} - f_L &= m_L\,(f_M - f_L) \\ R_{\hat{M}\cdot\sigma} - f_L &= m_{L\sigma}\,(f_M - f_L) \end{aligned}$$

By using an argument similar to that used to prove Proposition 1, we find that:

Proposition 10 *When* $0 < \alpha \leq 1$*, optimally* $R_{\hat{M}\cdot G} = f_M$ *and* $R_{\hat{M}\cdot B} = f_L$*. This implies that* $m_{LG} = 1$, $m_{LB} = 0$**.**

Consequently α and the low state audit probability solve

$$\min \alpha \left(\pi_{LG} c_m + c_a\right) + (1-\alpha)\,\pi_L m_L c_m$$

$$\begin{aligned} &\alpha\left\{(\pi_{HG} + \pi_{MG})\, f_M + (1 - \pi_{HG} - \pi_{MG})\, f_L - I - \pi_{LG} c_m - c_a\right\} + \\ &+ (1-\alpha)\left\{(\pi_H + \pi_M)\, R_{\hat{M}\cdot} + \pi_L\left(f_L - \frac{R_{\hat{M}\cdot} - f_L}{f_M - f_L} c_m\right) - I\right\} = 0 \end{aligned}$$

where the multiplier of α is the equivalent in a 3-state setting of EPC_σ:

$$EPC_{\sigma_3} \equiv \left\{(\pi_{HG} + \pi_{MG})\, f_M + (1 - \pi_{HG} - \pi_{MG})\, f_L - I - \pi_{LG} c_m - c_a\right\}.$$

As in the two state case there is a simple expression expressing the trade-off between the costs of signal acquisition and monitoring which tells us when the signal is worth it:

$$\begin{aligned} NC_{a_3} &\equiv (\pi_H + \pi_M)c_a - (\pi_{LB}(\pi_{HG} + \pi_{MG}) - \pi_{LG}(\pi_{HB} + \pi_{MB}))c_m \\ &\equiv (\pi_H + \pi_M)c_a - \rho_3 c_m \end{aligned}$$

Proposition 11 *When* $NC_{a_3} < 0$ *and the degree of correlation between signal and state is not too high, it is optimal to randomly or always gather information. The actual values of* $R_{\hat{M}\cdot}$, m_L *and* α *vary with the sign of* EPC_{σ_3}. *In particular:*

i. *If* $EPC_{\sigma_3} > 0$, *the unique optimum has* $m_L = 0$, $R_{\hat{M}\cdot} = f_L$,

$$\alpha = \underline{\alpha}_3 \equiv \frac{I - f_L}{(\pi_{HG} + \pi_{MG})(f_M - f_L) - \pi_{LG} c_m - c_a} < 1 \tag{44}$$

ii. *If* $EPC_{\sigma_3} = 0$, *it is optimal to always gather information* ($\alpha = 1$).

iii. *If* $EPC_{\sigma_3} < 0$, *the unique optimum has* $m_L = 1$, $R_{\hat{M}\cdot} = f_M$,

$$\alpha = \bar{\alpha}_3 \equiv \frac{(\pi_H + \pi_M) f_M + \pi_L f_L - I - \pi_L c_m}{(\pi_{HB} + \pi_{MB})(f_M - f_L) - \pi_{LB} c_m + c_a} < 1 \tag{45}$$

Applying a similar analysis to the one seen for the two-state cases, we can show that when NC_{a_3} is positive, then $\alpha = 0$, and when $NC_{a_3} = 0$, the principal is indifferent between collecting the signal or not. This allows us to sketch a diagram similar to Fig. 2, and with the same interpretation, with the $c_a - c_m$ space divided into areas according to the sign of EP_3, NC_{a_3} and EPC_{σ_3}.

So the tendency to generate deterministic monitoring also occurs with more than two states and arises because the income differences between states combined with the positive correlation of signal and state make it cheapest to concentrate monitoring on better signal states.

8 Discussion and extensions

In this section we analyse the role of some simplifying assumptions and discuss possible generalisations. To keep the notation simple we use two income states, but the discussion below applies also to three or more states cases.

8.1 Commitment

We have assumed commitment of the principal to the enforcement strategy written into the contract. As regards the audit strategy, it is well known that for a truthtelling contract to be credible, lenders must have some way of locking in to an agreed verification policy - some exogenous commitment device forcing the principal to audit low reports, even though she knows them to be truthful (Hart, 1995). While a deterministic audit policy may be enforceable if the report and the act of audit are public, this enforcement seems more problematic when the contract stipulates random audit, as it is impossible to infer the strategy actually used by the principal just from observation of whether an audit occurred (Khalil, 1997).[19] The only way round this is to insist that play of the random device requiring audit is public. In our setting, since optimally there are deterministic audits in the right hand branch when $NC_a < 0$, this problem is mitigated.[20]

The commitment problem to signal acquisition is rather different and its relevance depends on the sign of EPC_σ (17). If $EPC_\sigma = 0$, the principal cannot defect from the contracted deterministic strategy $\alpha = 1$, since any deviation is observable.

However, if $EPC_\sigma \neq 0$, the principal does have an incentive to cheat on a random signal acquisition strategy.[21] His return, defined in (16), is $\alpha EPC_\sigma + (1-\alpha)(PC_N - I) = 0$. If $EPC_\sigma > 0$, the contract has $\alpha < 1$ and $PC_N = f_L < I$, so he has an incentive to deviate setting $\alpha = 1$. Similarly if $EPC_\sigma < 0$, the contract has $0 < \alpha < 1$ and $PC_N = \pi_H f_H + \pi_L f_L - \pi_L c_m > I$, so he has an incentive to never collect the signal. Since play of the random device is unobservable, the only sequentially rational signal strategies are deterministic: $\alpha = 1$ when $EPC_\sigma > 0$ and $\alpha = 0$ when $EPC_\sigma < 0$. In these cases, lack of commitment will generate deterministic signal acquisition as well as deterministic audit. However, these contracts will give the agent a zero return in every signal-state outcome and a positive return to the principal.

From the above discussion, we derive the following proposition:

Proposition 12 *Under lack of commitment, the only sequentially rational signal acquisition strategies are deterministic.*

So, under no commitment, both audit and signal acquisition strategies are deterministic.

[19] The principal could cheat on the realisation of the random device, either by using a public device with the wrong odds or playing the device privately.

[20] Another way round the no commitment problem has the two parties engaging in some bargaining after the state is realised, e.g., collude to find an ex-post efficient repayment combined with no audit (Dewatripont, 1988), or noncooperatively trade-off the threat of audit against the threat of cheating on the report (Khalil, 1997; Khalil and Parigi, 1998). Knowing that some such process could occur, the contract would have to include a renegotiation-proof constraint, so that only those reporting and audit schemes that will actually be implemented subsequently are feasible. This leads to deterministic audit again (Krasa and Villamil, 2000).

[21] We have assumed that ex-ante competition between principals leads to a contract maximising the agent's expected payoff. However, ex-post, once the contract is agreed, the agent is locked in and the principal acts in his own interest.

8.2 Private information

Throughout the paper we have assumed that the signal is public. New issues arise if the signal is private information to the principal. In this case, depending on the timing of signal acquisition, the contract offer may have an informational content for the agent, becoming an informed principal problem (Maskin and Tirole (1990, 1992)). The main result in this literature is that with risk neutral parties the principal can do no better than publicly reveal her signal. However Cella (2008) shows that gains from hiding information are positive even with risk neutrality if there is correlation between the type of the principal and that of the agent, with two types of each. These results are derived in a context in which the signal is definitely acquired before the contract offer and it is a surplus-seeking principal who writes the contract. This is in contrast with our setup where the contract is written in the interests of the agent and the signal may be acquired either before or after the contract offer. Although these modelling differences do not allow us to immediately apply these results to our setting, an intuitive argument might run as follows. Suppose both parties know that a signal has been collected, but only the principal (who is rent-seeking) knows its outcome. Then the left and right hand branches of the game tree are still well defined. As far as the left hand branch of the game tree is concerned, a contract offer will convey information about the signal to the agent. If $\Pi_B \leq 0$ (30), the project profitability is negative after a bad signal and the principal will make a contract offer only after a good signal, thereby revealing it. The Maskin-Tirole result applies also in this setting. However, if $\Pi_B > 0$ (30), a contract offer will be made after any signal realisation. Because the returns are higher for the principal if a bad signal occurs ($R_{\hat{H} \cdot G|b4} < R_{\hat{H} \cdot B|b4}$ from Proposition 7), she may opportunistically hide the good signal and the Maskin-Tirole results may not apply. However, the possibility of competition among principals implies that competitors may also acquire information and undercut the principal's offer. Competitive offers would then reveal the signal. The principal could therefore decide to wait until after the contract to get the signal, given that competition has no force after the contract has been agreed. But a contract offer equal to the one with public signal may again induce the principal to hide information: she would now want to hide the bad signal, given that returns are higher after the good signal (recall $R_{\hat{H} \cdot G} = f_H > R_{\hat{H} \cdot B} = f_L$ from Proposition 1). Anticipating this, the agent would not accept a contract offer conditional on the signal and a pooling contract would presumably arise. The variety of the possible cases outlined above shows that the way private information would work out here needs a more thorough analysis.

8.3 Does it matter who gets the surplus?

Throughout the entire analysis we have taken the contract to allocate all the surplus to the agent so long as the principal breaks even ex-ante on the project. Besides being widely used in the CSV literature (Gale and Hellwig, 1985; Mookherjee and Png, 1989, among others), this assumption has social advantages as it minimises the deadweight loss of ensuring truthtelling. Indeed, without the signal ($\alpha = 0$), writing the contract to maximise the surplus of the principal subject to the participation constraint of the agent implies that, because of limited liability, the agent has to make a non-negative return in each state, i.e., his participation constraint must hold state by state. The principal can then raise the repayment required in a truthfully declared high state to the entire agents revenue by monitoring low reports with probability one ($m_N = 1$ and $R_{H \cdot N} = f_H$). This gives the principal the highest ex-ante return possible, but then the social cost of controlling the agent's incentives through monitoring is higher than it would be when the contract gives all the surplus to the agent. This can be seen by looking at the expected audit cost: $\pi_L c_m$, when the principal gets the surplus, and $\pi_L m_N c_m$ when the agent gets the surplus (Menichini and Simmons, 2012). Thus, having the principal writing the contract has an effect on the audit strategy, leading to deterministic audits already in the scenario with no signal, but at the social cost of a higher expected deadweight loss.

In a with-signal scenario, if the principal got the surplus, she would never choose to collect the signal. Again, since the agent's participation constraint must bind in each signal-state realisation, the principal can extract all revenue from the agent by monitoring each eventuality with probability one. But this implies that there is no return to using a signal to attain a finer classification of states and hence no need for a signal.

Concluding remarks

We have considered the contract problem in which an agent carries out a project generating private risky revenues which are subject to costly audit by a principal. The principal can choose at any time to get a costly and imperfectly informative signal about future revenues, whose realisation is freely available to the agent. The question is: how does the information in the signal control the debtors incentives? Should the principal get the signal and if so when? What impact does the signal have on the audit strategy and the structure of repayments in the contract?

Since the parties are risk-neutral, acquiring the signal has no risk sharing gain, but the reduction in uncertainty allows the principal to update and sharpen his beliefs about the prevalence of the incentive to cheat and may have incentive effects that reduce deadweight losses. There are three factors involved in the signal acquisition decision. Firstly, since collecting the signal reduces the uncertainty about future revenues, it allows the contract to be written with more precise separation of the possible types of agent. Secondly, since both the signal and audit are costly, it allows for a trade-off in control devices between paying to reduce uncertainty by getting a signal and paying for auditing. Thirdly, there are incentive effects of the signal which work indirectly through the principal's participation constraint and the agent's truthtelling constraint, that allow reduction in the deadweight loss of audits.

Sections 7 and 8 have discussed various possible extensions and open questions. One further issue concerns the possibility of delaying signal acquisition until after the agent has reported on the state. However, rather than representing a different timing of signal acquisition, this possibility represents a more sophisticated audit technology, given that, at the time at which the signal is acquired, the agent knows the true realisation of the state and is then unaffected by the outcome of the signal. In particular, it can be modelled as a sequential audit approach that first uses a cheap but imperfect screening device and then, if the results of that warrant it, goes on to a thorough more costly audit. We see this a fruitful topic for future research.

A Generic Properties of the Contract

(i) *the participation constraint must bind since otherwise the high state repayments can be reduced.*

(ii) and (iii) *the truthtelling (TT) constraints must bind and there must be maximum punishment* $R_{\hat{L}H\sigma} = f_H - c_a, R_{\hat{L}HN} = f_H$.

In all contracts the truthtelling constraint is

$$\begin{aligned} m_\sigma(R_{\hat{L}H\sigma} - R_{\hat{L}\cdot\sigma}) &\geq R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma} \geq 0 \qquad \sigma = G, B \\ m_N(R_{\hat{L}HN} - R_{\hat{L}\cdot N}) &\geq R_{\hat{H}\cdot N} - R_{\hat{L}\cdot N} \geq 0 \end{aligned}$$

The deadweight loss of monitoring is minimised by raising $R_{\hat{L}H\sigma}, R_{\hat{L}HN}$ to its maximum level of f_H (if $\alpha > 0$) and reducing m_σ, m_N until the TT holds with equality. In the left hand branch we must have $m_\sigma > 0$ since otherwise, for some σ, $R_{\hat{H}\cdot\sigma} = R_{\hat{L}\cdot\sigma} \leq f_L < I$ and there is then insufficient revenue to meet the investment cost.

(iv) *low state repayments in all contracts are set to give zero surplus to the firm.*

In the right hand branch (a similar argument applies to the left hand branch) after imposing binding truthtelling constraints and maximum punishment for cheating, the contract problem becomes

$$\begin{aligned} &\max\ \alpha \left\{ \textstyle\sum_\sigma \left[\pi_{H\sigma}\left(f_H - R_{\hat{H}\cdot\sigma}\right) + \pi_{L\sigma}\left(f_L - \left(1 - \frac{R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma}}{f_H - R_{\hat{L}\cdot\sigma}}\right) R_{\hat{L}\cdot\sigma} - \frac{R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma}}{f_H - R_{\hat{L}\cdot\sigma}} R_{\hat{L}L\sigma} \right) \right] \right\} \\ &+(1-\alpha)\left\{ \pi_H\left(f_H - R_{\hat{H}\cdot N}\right) + \pi_L \left[f_L - \left(1 - \frac{R_{\hat{H}\cdot N} - R_{\hat{L}\cdot N}}{f_H - R_{\hat{L}\cdot N}}\right) R_{\hat{L}\cdot N} - \frac{R_{\hat{H}\cdot N} - R_{\hat{L}\cdot N}}{f_H - R_{\hat{L}\cdot N}} R_{\hat{L}LN} \right] \right\} \\ \text{s.t.}\quad &\alpha \textstyle\sum_\sigma \left\{ \pi_{H\sigma} R_{\hat{H}\cdot\sigma} + \pi_{L\sigma}\left[\left(1 - \frac{R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma}}{f_H - R_{\hat{L}\cdot\sigma}}\right) R_{\hat{L}\cdot\sigma} + \frac{R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma}}{f_H - R_{\hat{L}\cdot\sigma}} (R_{\hat{L}L\sigma} - c_m) \right] - c_a \right\} \\ &+(1-\alpha)\left\{ \pi_H R_{\hat{H}\cdot N} + \pi_L \left(1 - \frac{R_{\hat{H}\cdot N} - R_{\hat{L}\cdot N}}{f_H - R_{\hat{L}\cdot N}}\right) R_{\hat{L}\cdot N} + \frac{R_{\hat{H}\cdot N} - R_{\hat{L}\cdot N}}{f_H - R_{\hat{L}\cdot N}} \left(R_{\hat{L}LN} - c_m\right) \right\} \geq I \end{aligned}$$

Forming a Lagrangian with multiplier λ, the FOC wrt $R_{\hat{H}\cdot\sigma}$ is (a very similar expression arises for $R_{\hat{H}\cdot N}$):

$\frac{\partial \mathcal{L}}{\partial R_{\hat{H}\cdot\sigma}} : \alpha(\lambda - 1)\left(\pi_{H\sigma} - \frac{\pi_{L\sigma}\left(R_{\hat{L}\cdot\sigma} - R_{\hat{L}L\sigma}\right)}{f_H - R_{\hat{L}\cdot\sigma}} \right) - \alpha\lambda \frac{\pi_{L\sigma} c_m}{f_H - R_{\hat{L}\cdot\sigma}} \geq 0, R_{\hat{H}\cdot\sigma} \leq f_H.$

At the optimum $\lambda > 1$. If not, the FOCs wrt $R_{\hat{H}\cdot\sigma}, R_{\hat{H}\cdot N}$ are all negative, so all repayments are at most f_L. But this would yield insufficient revenue to the principal.

Using the FOC wrt $R_{\hat{L}\cdot\sigma}$:

$\frac{\partial \mathcal{L}}{\partial R_{\hat{L}\cdot\sigma}} : \frac{\alpha \pi_{L\sigma}}{\left(f_H - R_{\hat{L}\cdot\sigma}\right)^2} \left\{ (\lambda - 1)\left(f_H - R_{\hat{H}\cdot\sigma}\right)\left(f_H - R_{\hat{L}L\sigma}\right) + \lambda c_m \left(f_H - R_{\hat{H}\cdot\sigma}\right) \right\} \geq 0, R_{\hat{L}\cdot\sigma} \leq f_L.$

This is positive and hence $R_{\hat{L}\cdot\sigma} = f_L$. A similar argument holds for $R_{\hat{L}\cdot N}$.

The FOC wrt $R_{\hat{L}L\sigma}$ is (a similar expression arises for $R_{\hat{L}LN}$): $\frac{\partial \mathcal{L}}{\partial R_{\hat{L}L\sigma}} : \alpha(\lambda - 1)\pi_{L\sigma} \frac{R_{\hat{H}\cdot\sigma} - R_{\hat{L}\cdot\sigma}}{f_H - R_{\hat{L}\cdot\sigma}} \geq 0, R_{\hat{L}L\sigma} \leq f_L$.

This is positive (also for no signal state N) and hence optimally $R_{\hat{L}L\sigma} = R_{\hat{L}LN} = f_L$. Hence low state audited and nonaudited repayments are set to give zero rent to the firm.

Using $R_{\hat{L}\cdot\sigma} = R_{\hat{L}L\sigma} = R_{\hat{L}\cdot N} = R_{\hat{L}LN} = f_L$, we obtain programme $\mathcal{P}_{\mathcal{EC}}$.

B Signal after the contract

The rest of the proof proceeds as follows. We have four variables to choose m_σ, m_N, α. We first show that one set of optimal solution values for m_G, m_B are $m_G = 1, m_B = 0$ for any admissible values of $0 \leq \alpha, m_N \leq 1$. In fact if $c_a > 0$ these are the only optimal values when $\alpha > 0$. When $\alpha = 0$ they are admissible but so are any others (in fact m_G, m_B are then irrelevant). When $c_a = 0$ there are also multiple combinations of optimal α, m_G or α, m_B depending on the probability distribution of revenues. Using $m_G = 1, m_B = 0$ then gives information about the optimal values of α, m_N (through the trade-off between α and m_N in the objective and constraint).

Proposition 1 shows that if optimally $\alpha > 0$ then $m_G = 1, m_B = 0$. Proposition 3 shows that if $NC_a < 0$ and $c_a > 0$ then optimally $\alpha > 0$ and, depending on the probability distribution of revenues, either $m_N = 0, m_N = 1$ or $\alpha = 1$ and m_N is irrelevant. Proposition 2 shows that if $NC_a > 0$ and $c_a > 0$ then optimally $\alpha = 0$ and m_N assumes the value of the no-signal contract. Proposition 4 takes the watershed case $NC_a = 0$ and $c_a > 0$ in which case the no-signal type monitoring with $\alpha = 0$ and a random signal acquisition policy are indifferent. Finally Proposition 5 analyses the $c_a = 0$ case and shows that as well as the $\alpha > 0, m_G = 1, m_B = 0$ case, an infinite number of other optimal solutions exist which involve mixes of $\alpha, m_G > 0$ or mixes of $\alpha, m_G > 0$ according the distribution of revenue between states.

Proof of Proposition 1. Suppose that $0 < \alpha < 1$ and $c_a > 0$.

1. An optimum cannot have all $m_\sigma = m_N = 1, \sigma = G, B$. If it did, the participation constraint (12) would be slack ($Ef - I - \pi_L c_m - c_a > 0$ by assumption) and a rent would be needlessly left to the principal. Similarly, it cannot have all $m_\sigma = m_N = 0, \sigma = G, B$, since then the expected repayments would not meet the reservation level of the principal ($f_L - I < 0$). Thus at least one audit probability must be positive.

2. Suppose that there are $m_\sigma, m_N > 0, \sigma = G, B$, and $m_G < 1$ such that the participation constraint (12) holds. This is not an optimum because it is possible to lower m_B (i.e. lower $R_{\hat{H}\cdot B}$) and increase m_G (i.e. increase $R_{\hat{H}\cdot G}$) while reducing the expected cost (14) and slackening the participation constraint (12). Indeed because $\pi_{HG} > \pi_{HB}$ the increase in m_G ($R_{\hat{H}\cdot G}$) and reduction in m_B ($R_{\hat{H}\cdot B}$) allows an increase in the expected revenues after a good signal larger than the fall in revenues after the bad signal. Similarly this shift reduces the with signal expected audit cost since $\pi_{LG} < \pi_{LB}$. This can be seen in the following expression:

$$\frac{1}{f_H - f_L}\left[\left(\pi_{HG}\left(f_H - f_L\right) - \pi_{LG}c_m\right)dR_{\hat{H}\cdot G} + \left(\pi_{HB}\left(f_H - f_L\right) - \pi_{LB}c_m\right)dR_{\hat{H}\cdot B}\right].$$

Since $\left(\pi_{HG}\left(f_H - f_L\right) - \pi_{LG}c_m\right) - \left(\pi_{HB}\left(f_H - f_L\right) - \pi_{LB}c_m\right) > 0$, the participation constraint is slackened by an increase in m_G matched by an equal reduction in m_B. This allows a further reduction in m_B and thus a reduction of the objective function. So an optimum cannot have $m_G < 1$ and $m_B > 0$ and must at least entail one of $m_G = 1$ or $m_B = 0$.

3. Suppose that $m_G = 1$ but $1 > m_B > 0$. Then if $m_N < 1$, it is possible to vary m_N and m_B keeping the participation constraint constant

$$dm_N = -\frac{\alpha\left(\pi_{HB}(f_H - f_L) - \pi_{LB}c_m\right)}{(1-\alpha)\left(\pi_H(f_H - f_L) - \pi_L c_m\right)} dm_B < 0 \tag{46}$$

Thus, m_B and m_N must vary in opposite directions to satisfy the participation constraint. The effect of such variations on the objective function (14) is $dEC = \alpha c_m\left[-\pi_{LG} dm_N + \pi_{LB}(dm_B - dm_N)\right] + \pi_L c_m dm_N$, which, using (46), becomes $dEC = \alpha c_m \frac{\rho(f_H - f_L)}{\pi_H(f_H - f_L) - \pi_L c_m} dm_B$. Since the coefficient of dm_B is positive, it is best to reduce m_B and raise $m_{N.}$ as far as possible.

   (a) One possibility is to hit $m_B = 0$ with $m_N > 0$. This proves the result.

   (b) Alternatively $m_N = 1$ with $m_B > 0$ still. In this last case the optimisation problem becomes:

$$\min \alpha\left[c_a - \pi_{LB}c_m(1 - m_B)\right] + \pi_L c_m$$
$$\text{s.t. } -\alpha\left\{(1-m_B)\left[\pi_{HB}(f_H - f_L) - \pi_{LB}c_m\right] + c_a\right\} + (\pi_H f_H + \pi_L f_L - \pi_L c_m) = I$$

   It is then possible to vary α and m_B keeping the participation constraint constant

$$d\alpha = \frac{\alpha\left(\pi_{HB}(f_H - f_L) - \pi_{LB}c_m\right)}{(1-m_B)\left[\pi_{HB}(f_H - f_L) - \pi_{LB}c_m\right] + c_a} dm_B > 0 \tag{47}$$

   The effect on the objective function (14) is $dEC = \left[c_a - \pi_{LB}c_m(1-m_B)\right]d\alpha + \alpha\pi_{LB}c_m dm_B$, whence, using (47) gives $dEC = \frac{\alpha c_a \pi_{HB}(f_H - f_L)}{(1-m_B)\left[\pi_{HB}(f_H - f_L) - \pi_{LB}c_m\right] + c_a} dm_B > 0$. It is possible to reduce m_B to zero whilst keeping $\alpha > 0$ since, with $m_G = m_N = 1$, $m_B = 0$, the participation constraint becomes $\alpha EPC_\sigma + (1-\alpha)(\pi_H f_H + \pi_L f_L - I - \pi_L c_m) \geq 0$. Since $(\pi_H f_H + \pi_L f_L - I - \pi_L c_m) > 0$, this can be satisfied with $\alpha > 0$.

4. Alternatively suppose the optimum has $m_B = 0$ but $0 < m_G < 1$ and $m_N > 0$. It is then possible to reduce m_N, raise m_G so as to keep the participation constraint constant: $dm_N = -\frac{\alpha\{\pi_{HG}(f_H - f_L) - \pi_{LG}c_m\}}{(1-\alpha)\{\pi_H(f_H - f_L) - \pi_L c_m\}} dm_G$. The change in objective (14) is $dEC = \alpha c_m \frac{-\rho(f_H - f_L)}{\pi_H(f_H - f_L) - \pi_L c_m} dm_G < 0$. So it is best to raise m_G to maximum and reduce m_N.

   (a) One possibility is that we hit $m_G = 1$ with $m_N > 0$, which proves the result.

   (b) The other possible case has $m_N = 0$ with $m_G < 1$ still. In this last case the optimisation problem becomes

$$\min \alpha\left(c_a + \pi_{LG} m_G c_m\right)$$
$$\text{s.t. } \alpha\left\{\left(\pi_{HG}(f_H - f_L) - \pi_{LG}c_m\right)m_G - c_a\right\} = I - f_L.$$

   Varying m_G and α so as to keep the participation constraint constant gives $dm_G = -\frac{(\pi_{HG}(f_H - f_L) - \pi_{LG}c_m)m_G - c_a}{\alpha[\pi_{HG}(f_H - f_L) - \pi_{LG}c_m]} d\alpha$. Using the resulting dm_G in the objective function $dEC = (c_a + c_m\pi_{LG}m_G)\, d\alpha + \alpha c_m \pi_{LG} dm_G$ gives $dEC = \frac{\pi_{HG}(f_H - f_L)c_a}{\pi_{HG}(f_H - f_L) - \pi_{LG}c_m} d\alpha > 0$. So it is best to reduce α and raise $m_{G.}$ as far as possible. In fact we can attain $m_G = 1$ with $\alpha > 0$.

Hence the optimum must involve $m_G = 1, m_B = 0$. ■

Proposition 1 is valid if $\alpha > 0$. If $\alpha = 0, m_\sigma$ have no impact on participation constraint or objective and so without loss of generality we can also set them to these values if $\alpha = 0$.

To prove Propositions 2 to 5, we use the following lemma.

Lemma 1 *If $0 \leq \alpha < 1$ and $0 < m_N < 1$ are optimal, if $EPC_\sigma \neq 0$, a simultaneous variation in α and m_N keeping the participation constraint constant changes expected enforcement cost according to $dEC = \frac{(\pi_H c_a - \rho c_m)(f_H - f_L)}{[\pi_H(f_H - f_L) - \pi_L c_m]} d\alpha$, whose sign depends on the sign of $\pi_H c_a - \rho c_m \equiv NC_a$.*

Proof. Consider Programme $\mathcal{P}'_{\mathcal{EC}}$. Suppose $0 \leq \alpha < 1$ and $0 < m_N < 1$ are optimal. When $EPC_\sigma \neq 0$ simultaneous variation in α, m_N to keep the participation constraint constant must satisfy

$$\frac{dm_N}{d\alpha}|_{PC} = -\frac{EPC_\sigma - (PC_N - I)}{(1-\alpha)\left[\pi_H(f_H - f_L) - \pi_L c_m\right]} \tag{48}$$

While the sign of the denominator is certainly positive, the sign of the numerator is ambiguous. However, it can be inferred from the sign of EPC_σ. For $\alpha \in (0,1)$, to satisfy (16) with equality, $PC_N - I \lesseqgtr 0$ if $EPC_\sigma \gtreqless 0$. Thus, if:

- $EPC_\sigma > 0$, $\frac{dm_N}{d\alpha}|_{PC} < 0$;
- $EPC_\sigma < 0$, $\frac{dm_N}{d\alpha}|_{PC} > 0$.

The simultaneous variation in m_N, α changes the expected enforcement cost by $(c_a + \pi_{LG} c_m - \pi_L m_N c_m)\, d\alpha + (1-\alpha)\, \pi_L c_m dm_N$, i.e., using 48,

$$(c_a + \pi_{LG} c_m - \pi_L m_N c_m)\, d\alpha + (1-\alpha)\, \pi_L c_m \left(-\frac{EPC_\sigma - (PC_N - I)}{(1-\alpha)\left[\pi_H(f_H - f_L) - \pi_L c_m\right]} \right) d\alpha$$

which reduces to

$$dEC = \frac{(\pi_H c_a - \rho c_m)(f_H - f_L)}{[\pi_H(f_H - f_L) - \pi_L c_m]} d\alpha.^{22} \tag{49}$$

- If $EPC_\sigma = 0$, $PC = (1-\alpha)(m_N(\pi_H(f_H - f_L) - \pi_L c_m) - I + f_L)$ and given $\alpha < 1$ we must have $m_N(\pi_H(f_H - f_L) - \pi_L c_m) - I + f_L = 0$. In this case no marginal compensation between m_N, α is possible.

■

Proof of Proposition 2. If $NC_a > 0$, from (49) enforcement costs are minimised by reducing α. However, the change in m_N following a change in α necessary to preserve the participation constraint depends on the sign of EPC_σ. Three cases can arise:

1. If $EPC_\sigma > 0$, $\frac{dm_N}{d\alpha}|_{PC} < 0$. Thus m_N and α vary in opposite directions to satisfy the participation constraint. So, to minimise the enforcement cost, it is best to raise m_N

[22] A similar procedure applies to work out dEC/dm_N.

and decrease α. However, an optimum cannot have $m_N = 1$ and $\alpha > 0$, because the participation constraint would be slack $(\alpha EPC_\sigma + (1-\alpha)(PC_N - I) = 0$. If $m_N = 1$, $PC_N - I = \pi_H f_H + \pi_L f_L - I - \pi_L c_m > 0$. Given $EPC_\sigma > 0$, $PC > 0)$. The firm can then reduce m_N until the participation constraint binds, giving $PC_N - I < 0$ (given that $EPC_\sigma > 0$). But because $\frac{dEC}{dm_N} = -\frac{(1-\alpha)(f_H - f_L)(\pi_H c_a - \rho c_m)}{EPC_\sigma - (PC_N - I)} < 0$, it is best to raise m_N and lower α. However m_N only has to rise to balance PC and we know that the no-signal solution with $\alpha = 0$ is achievable with $m_N < 1$. If α were positive but small there would be further gains to be realised. Thus optimally $\alpha = 0$, with $R_{\hat{H}\cdot N} = R_{\hat{H}\cdot N}|_{\alpha=0}$ (21).

2. If $EPC_\sigma < 0$, $\frac{dm_N}{d\alpha}|_{PC} > 0$. Now α, m_N must both fall to balance PC. Again we know that we can achieve $\alpha = 0$ with $PC = 0$ at the no-signal solution. And if we left $\alpha > 0$ there would still be gains to be realised from reducing both α, m_N. Thus the solution must have $m_N > 0$, $R_{\hat{H}\cdot N} = R_{\hat{H}\cdot N}|_{\alpha=0}$ (21), and $\alpha = 0$.

3. If $EPC_\sigma = 0$, $PC = (1-\alpha)(PC_N - I)$. One possibility is $\alpha = 1$, in which case m_N is immaterial, the other is $m_N P_N - I + f_L = 0$ in which case, since $c_a > 0$, it is best to set $\alpha = 0$. Of these two possibilities, if $NC_a > 0$ it is best to set $\alpha = 0$.

In either case enforcement costs are $EC_{\alpha=0} = \pi_L m_N c_m$ (22). ■

Proof of Proposition 3. If $NC_a < 0$, from (49) $\frac{dEC}{d\alpha} < 0$: enforcement costs are minimised by raising α. However, the change in m_N following a change in α necessary to preserve the participation constraint depends on the sign of EPC_σ. Three cases can arise:

1. If $EPC_\sigma > 0$, from (48) $\frac{dm_N}{d\alpha}|_{PC} < 0$. Thus m_N and α vary in opposite directions to satisfy the participation constraint. To minimise the enforcement cost, it is best to reduce m_N as far as possible resulting in $m_N = 0$ $(R_{\hat{H}\cdot N} = f_L)$ and raise α just far enough to satisfy the participation constraint to $\underline{\alpha}$. Setting $\alpha = 1$ (and m_N interior) is not optimal as it would give a slack participation constraint.

2. If $EPC_\sigma = 0$, since $NC_a < 0$ now it is best to set $\alpha = 1$ and then m_N is immaterial.

   If $EPC_\sigma = 0$ there are two possible ways in which the participation constraint can hold: either $\alpha = 1$ or m_N is selected so that $PC_N = I$. This in turn gives $m_N|_{\alpha=0} = \frac{I - f_L}{\pi_H(f_H - f_L) - \pi_L c_m}$ ($m_N|_{\alpha=0}$ is defined in (20)) If the second route is taken, there is still a free choice of α. However, when $m_N = m_N|_{\alpha=0}$, $\frac{dEC}{d\alpha} = \frac{(\pi_H c_a - \rho c_m)(f_H - f_L) - \pi_{LG} c_a c_m}{\pi_H(f_H - f_L) - \pi_L c_m} < 0$, and so the best choice of α, given this, is $\alpha = 0$. Thus, if $EPC_\sigma = 0$ we can reach $EC_{\alpha=1} = c_a + \pi_{LG} c_m$ via $\alpha = 1$ or $EC_{\alpha=0} = \frac{\pi_L(I - f_L)c_m}{\pi_H(f_H - f_L) - \pi_L c_m}$ with $\alpha = 0$. Comparing $EC_{\alpha=1}$ with $EC_{\alpha=0}$ we obtain $(\pi_H c_a - \rho c_m)(f_H - f_L) < 0$: the optimum when $EPC_\sigma = 0$ is to set $\alpha = 1$ and have full information gathering.

3. If $EPC_\sigma < 0$, $\frac{dm_N}{d\alpha}|_{PC} > 0$. Thus m_N and α vary in the same direction to satisfy the participation constraint and α has to be increased to minimise the enforcement cost. Setting

$\alpha = 1$ is not optimal as it would violate participation constraint. Thus m_N must be raised as far as possible, giving $m_N = 1$ and α must be raised sufficiently to satisfy the participation constraint: $\alpha = \bar{\alpha}$ (26).

To see why policing involves concentrated audit with $m_G = 1$ and α interior, consider the enforcement costs when m_G is interior: $EC = \alpha \left(c_a + \pi_{LG} m_G c_m \right)$. Proportional and opposite changes in m_G or α have identical expected audit cost effects, but in addition the expected signal acquisition cost is αc_a. Thus, the extra cost of a proportional increase in the audit probability m_G is more than offset by the saving induced by a reduction in signal acquisition frequency α. It is then optimal to increase the audit probability all the way until $m_G = 1$ and reduce the chance of signal acquisition until the participation constraint binds. Thus, it is the direct cost of collecting the signal that ultimately makes the signal acquisition strategy more expensive than auditing and leads to deterministic audits.

If the project is not ex-ante profitable after a signal ($EPC_\sigma < 0$), then high state repayments in no signal states must be increased to satisfy the participation constraint, thus necessitating audit of no signal states too. Signal contingent deterministic monitoring is still optimal ($m_G = 1, m_B = 0$), but now signal acquisition has effects on the relative expected audit costs of good signal and no signal low reports . The optimal combination α, m_N minimises $EC = \alpha \left(c_a + \pi_{LG} c_m \right) + (1 - \alpha) \pi_L m_N c_m$. In trading off α and m_N, each has a specific cost $(c_a + \pi_{LG} c_m)$ for α and $m_N \pi_L c_m$ for m_N, as well as a common saving of $\alpha \pi_L m_N c_m$, since, when there is a signal, m_N is not applied. From the fact that $NC_a < 0$ and the signal is relatively cheap, it turns out that the cost saving from reducing α exceeds that from reducing m_N and it is optimal to increase the audit probability all the way until $m_N = 1$ and use α just to satisfy the participation constraint. ■

Proof of Proposition 4. If $NC_a = 0$ then $dEC/d\alpha = 0$ so simultaneous variation in α and m_N (R_{HN}) which ensure that the participation constraint of the principal continues to hold have a zero effect on policing cost. It follows that any combination of values α, m_N in the unit interval which satisfy the participation constraint will generate the same optimal policing cost and there is an infinity of solutions. The permissible ranges of α, m_N are derived from the participation constraint (16) and depend on the sign of EPC_σ.

1. If $EPC_\sigma > 0$, from (48) $\frac{dm_N}{d\alpha}|_{PC} < 0$, so m_N and α must move in opposite direction to satisfy the participation constraint. To ensure that α, m_N remain in the unit interval the range of values of α is $\alpha \in [0, \underline{\alpha}]$ with a corresponding range of values of $m_N \in [0, m_N|_{\alpha=0}]$ ($R_{HN} \in [f_L, R_{HN}|_{\alpha=0}]$), where $\underline{\alpha}$ and $m_N|_{\alpha=0}$ are defined in (23) and (20) respectively.

2. If $EPC_\sigma < 0$, α and m_N must move in the same direction along the PC (from (48) $\frac{dm_N}{d\alpha}|_{PC} > 0$). To ensure that α, m_N remain in the unit interval the range of values of α is $\alpha \in [0, \bar{\alpha}]$,with $\bar{\alpha}$ as defined in (26), with a corresponding range of values of $m_N \in [m_N|_{\alpha=0}, 1]$ ($R_{HN} \in [R_{HN}|_{\alpha=0}, f_H]$).

3. If $EPC_\sigma = 0$, the multiplier of α in PC is zero and the PC is satisfied either by setting $m_N = m_N|_{\alpha=0}$ (in which case $\alpha = 0$) of by setting $\alpha = 1$, in which case m_N is undefined.

In either case enforcement costs are equal to those obtained when $\alpha = 0$ (22). ■

Proof of Proposition 5. From Proposition 1 we know that an optimum cannot have $m_G < 1$ and $m_B > 0$ and must at least entail one of $m_G = 1$ or $m_B = 0$.

1. Suppose that $m_G = 1$ but $1 > m_B > 0$. Then if $m_N < 1$, as in the case of Proposition 1 with $c_a > 0$, it is best to reduce m_B and raise $m_{N.}$ as far as possible. There are two possibilities:

   (a) One has $m_N = 1$ with $m_B > 0$. This case has been dealt with in point 3 of the Proof of Proposition 1. With $c_a = 0$, the trade-off between m_B and α is given by
   $$dEC = \alpha \left[-\pi_{LB} c_m (1 - m_B) \left(\frac{(\pi_{HB} (f_H - f_L) - \pi_{LB} c_m)}{(1 - m_B) [\pi_{HB} (f_H - f_L) - \pi_{LB} c_m]} \right) + \pi_{LB} c_m \right] dm_B = 0$$
   and the firm is indifferent between monitoring after a bad signal or acquiring the signal. The solution has $\alpha \in \left[\min \left\{ \frac{\pi_H f_H + \pi_L f_L - I - \pi_L c_m}{\pi_{HB}(f_H - f_L) - \pi_{LB} c_m}, 1 \right\}, 1 \right]$ and $m_B \in \left[0, \max \left\{ 0, -\frac{EPC_\sigma}{\pi_{HB}(f_H - f_L) - \pi_{LB} c_m} \right\} \right]$.

   (b) Alternatively $m_B = 0$, $m_N > 0$. This case is the analogous to the one dealt with in Lemma 1 with $NC_a < 0$ always, given $c_a = 0$. The solution depends on the sign of EPC_σ and is described in Proposition 3.

2. Alternatively suppose the optimum has $m_B = 0$ but $0 < m_G < 1$ and $m_N > 0$. As in the case of Proposition 1, it is desirable to reduce m_N and raise m_G as far as possible. Then there are two possibilities:

   (a) $m_N = 0$, $m_G < 1$. This case has been dealt with in point 4 of the Proof of Proposition 1. With $c_a = 0$, the trade-off between m_G and α becomes $dEC = (c_m \pi_{LG} m_G) d\alpha + c_m \pi_{LG} (-m_G d\alpha) = 0$. The firm is indifferent between monitoring after a good signal or acquiring the signal. The solution has $\alpha \in \left[1, \min \left\{ \frac{I - f_L}{\pi_{HG}(f_H - f_L) - \pi_{LG} c_m}, 1 \right\} \right]$ and $m_G \in \left[\min \left\{ \frac{I - f_L}{\pi_{HG}(f_H - f_L) - \pi_{LG} c_m}, 1 \right\}, 1 \right]$. The enforcement costs are $EC = \alpha \pi_{LG} m_G c_m = \pi_{LG} \frac{I - f_L}{\pi_{HG}(f_H - f_L) - \pi_{LG} c_m} c_m$.

   (b) $m_G = 1$, $m_N > 0$. This case is the analogous of the one dealt with in Lemma 1 with $NC_a < 0$ always, given $c_a = 0$. The solution depends on the sign of EPC_σ and is described in Proposition 3.

Thus, as in the case in which $c_a > 0$, the solution depends on the sign of EPC_σ. When positive, starting from the equilibrium with costly signal acquisition ($m_G = 1$, $\alpha = \underline{\alpha}$), an increase in α can be exactly compensated by a proportional decrease in m_G, thus keeping the expected deadweight loss of audit constant. This gives rise to a continuum of possible equilibria with $\underline{\alpha} \leq \alpha \leq 1$ and audit after a good signal at a rate $m'_G \geq m_{G|\alpha=1} = \frac{I - f_L}{\pi_{HG}(f_H - f_L) - \pi_{LG} c_m}$ (rising to $m'_G = 1$ when

$\alpha = \underline{\alpha}$, where $\underline{\alpha}$ is always set to the level that satisfies the participation constraint). If $EPC_\sigma = 0$, the participation constraint (16) can be satisfied either by always getting the signal or by audit of no signal states. Since the signal is now free, a fortiori it is cheaper to set $\alpha = 1$. If $EPC_\sigma < 0$, α, m_N and m_B trade-off against each other. We know that $m_N = 1$ from Proposition 3 and the enforcement cost is $EC = [\pi_L - \alpha\pi_{LB}(1 - m_B)] c_m$. Any simultaneous increase in α and m_B gives $dEC = \pi_{LB}(\alpha dm_B - (1 - m_B)d\alpha)$. So, choosing $d\alpha = \alpha/(1 - m_B)\, dm_B$ keeps enforcement cost constant. This gives rise to a continuum of possible equilibria with $\bar{\alpha} \le \alpha \le 1$ and audit after a bad signal at a rate $0 \le m'_B \le m_{B|\alpha=1} = -\frac{EPC_\sigma}{\pi_{HB}(f_H - f_L) - \pi_{LB} c_m}$. ∎

Proof of Proposition 6. Let c^*_m, c^*_a solve $EP = EPC_\sigma = 0$. Note that in the c_a, c_m plane EP falls faster than EPC_σ. Similarly let $\widehat{c}_m, \widehat{c}_a$ solve $EPC_\sigma = NC_a$. $c^*_a = \widehat{c}_a$, $c^*_m = \widehat{c}_m$ if $\rho = \rho^*$ and if $\rho > \rho^*, c^*_a > \widehat{c}_a$ and $c^*_m < \widehat{c}_m$.

Feasibility requires c_a, c_m are below EP. If $c^*_a < \widehat{c}_a$ there exist feasible pairs c_a, c_m at which $EPC_\sigma \le 0$ with $c_a < \widehat{c}_a, c_m > \widehat{c}_m$ but $NC_a < 0$. At any such point the optimal signal strategy is MS_b if $EPC_\sigma < 0$ or $\alpha = 1$ if $EPC_\sigma = 0$. And at points c_a, c_m with $EPC_\sigma > 0$ it is MS_a

The difference $c^*_a - \widehat{c}_a$ (which has the opposite sign of $c^*_m - \widehat{c}_m$) is equal to $\rho\pi_{HG}(f_H - f_L) - \pi_H\pi_{LB}(I - f_L)$. It is increasing in ρ and zero when $\rho = \rho^* \equiv \frac{\pi_H\pi_{LB}(I-f_L)}{\pi_{HG}(f_H-f_L)}$. Hence if $\rho \le \rho^*$ we have $c^*_a < \widehat{c}_a$ and all three parts of proposition 3 apply.

Conversely if $\rho > \rho^*, c^*_a > \widehat{c}_a$ and $c^*_m < \widehat{c}_m$. Now the only feasible points $c_a c_m$ below EP and with $NC_a \le 0$ must entail $EPC_\sigma > 0$ and hence whenever $NC_a \le 0$, the signal strategy can only be MS_a if it acquired at all. ∎

Comparative statics: Signal strategy as a function of acquisition and audit cost. We first plot α against c_a and c_m. When $NC_a < 0$ and $EPC_\sigma \ge 0$, $\frac{\partial\underline{\alpha}}{\partial c_m} = \frac{\pi_{LG}(I-f_L)}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^2} > 0$, $\frac{\partial^2\underline{\alpha}}{\partial c_m^2} = \frac{2\pi_{LG}^2\alpha}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^2} > 0$.

When $EPC_\sigma < 0$, $\frac{\partial\bar{\alpha}}{\partial c_m} = -\frac{\pi_L-\alpha\pi_{LB}}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^2} < 0$, $\frac{\partial^2\bar{\alpha}}{\partial c_m^2} = -\frac{2\pi_{LB}(\pi_L-\alpha\pi_{LB})}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^2} < 0$. At its maximum, the slope of $\alpha(c_m)$ changes from $\frac{\partial\alpha}{\partial c_m} = \frac{\pi_{LG}}{I-f_L}$ to $\frac{\partial\alpha}{\partial c_m} = \frac{\pi_{LG}^2}{\rho(f_H-f_L)-\pi_{LB}(I-f_L)-\pi_L c_a} < 0$.[23]

When $EPC_\sigma \ge 0$, $\frac{\partial\underline{\alpha}}{\partial c_a} = \frac{I-f_L}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^2} > 0$, $\frac{\partial^2\underline{\alpha}}{\partial c_a^2} = \frac{2(I-f_L)}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^3} > 0$.

When $EPC_\sigma < 0$, $\frac{\partial\bar{\alpha}}{\partial c_a} = -\frac{Ef-I-\pi_L c_m}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^2} < 0$, $\frac{\partial^2\bar{\alpha}}{\partial c_a^2} = \frac{2(Ef-I-\pi_L c_m)}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^3} > 0$.

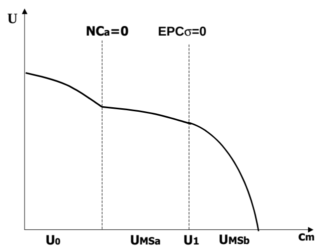
Since utility is given by $U = \pi_H f_H + \pi_L f_L - I - EC$, where EC is the relevant cost function, we can easily plot U against c_a and c_m. For $c_a > 0$ and increasing c_m, optimally $\alpha = 0$ and from 22 utility is given by $U_{\alpha=0} = \frac{\pi_H(f_H-f_L)(\pi_H f_H+\pi_L f_L-I-\pi_L c_m)}{\pi_H(f_H-f_L)-\pi_L c_m}$. Then $\frac{\partial U_{\alpha=0}}{\partial c_m} = -\frac{\pi_H\pi_L(f_H-f_L)(I-f_L)}{(\pi_H(f_H-f_L)-\pi_L c_m)^2} < 0$; $\frac{\partial U^2_{\alpha=0}}{\partial c_m^2} = -\frac{2\pi_H\pi_L^2(f_H-f_L)(I-f_L)}{(\pi_H(f_H-f_L)-\pi_L c_m)^3} < 0$.

For higher c_m, optimally $\alpha > 0$ and from 24 utility is given by $U^a_{MS} = \frac{(f_H-f_L)\{\pi_{HB}(I-f_L)+\pi_H EPC_\sigma\}}{\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a}$ and $\frac{\partial U^a_{MS}}{\partial c_m} = -\frac{\pi_{LG}\pi_{HG}(f_H-f_L)(I-f_L)}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^2} < 0$; $\frac{\partial U^{a^2}_{MS}}{\partial c_m^2} = -\frac{2\pi_{LG}^2\pi_{HG}(f_H-f_L)(I-f_L)}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m-c_a)^2} < 0$.

[23] For this to be negative, we need $EP_B = \pi_{HB}(f_H - f_L) - \pi_{LB}c_m > 0$ at the value of c_m at which $EPC_\sigma = 0$. This follows from assumption 1 and $EP = EP_B + EPC_\sigma$.

Last, for sufficiently high c_m, optimally still $\alpha > 0$, but from 27 utility is given by $U^b_{MS} = \frac{(Ef-I-\pi_L c_m)\pi_{HB}(f_H-f_L)}{\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a}$ and $\frac{\partial U^b_{MS}}{\partial c_m} = \frac{\pi_{HB}(f_H-f_L)\{\pi_{LB}EPC_\sigma-\pi_{LG}[\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a]\}}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^2} < 0$; $\frac{\partial U^{b^2}_{MS}}{\partial c^2_m} = -\frac{2\pi_{LB}\pi_{HB}(f_H-f_L)\{\pi_{LB}EPC_\sigma-\pi_{LG}[\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a]\}}{(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m+c_a)^3} < 0$.

The pattern of utility for varying c_a and c_m is depicted in figures 8 and 9:



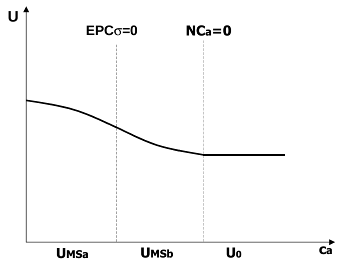


Fig. 8. $U(c_m)$ for $c_a > 0$. Fig. 9. $U(c_a)$ for $c_m > 0$.

C Signal before the contract

Proof of Proposition 7. When $\Pi_B > 0$, using programme $\mathcal{P}'_{b4}$, solving the participation constraint for $R_{\hat{H}\cdot\sigma}$ (32) and substituting out both in m_σ and in the utility function U_σ gives $m_{\sigma|b4}$ and EU_{b4} reported in Proposition 7 (31 and 33) respectively). Notice that $m_{G|b4} - m_{B|b4} = \frac{-(I-f_L)\rho(f_H-f_L+c_m)}{(\pi_{HG}(f_H-f_L)-\pi_{LG}c_m)(\pi_{HB}(f_H-f_L)-\pi_{LB}c_m)} < 0$. When $\Pi_B \leq 0$, the principal knows he will make zero profit at best from proceeding after a bad signal and generally will abandon the project. The outcomes after a good signal are not affected, so the expected return to the agent arises only from the good signal outcome (34). ■

D Globally optimal signal strategy

Proof of Proposition 8. Let R_{b4} be the optimal repayments in the left hand (LH) branch (signal before the contract), α, R_{after} the optimal signal strategy and repayments in the right hand (RH) branch (signal after the contract).

R_{b4} satisfies the truthtelling constraint (TT), the participation constraint (PC) in each signal state $E_{s|\sigma}PC_\sigma = 0, \sigma = G, B$, and the relevant limited liability (LL) constraints (10).

The maximum payoff in LH branch is $EU_{b4} = \sum_\sigma \pi_\sigma E_{s|\sigma}U_\sigma(R_{\sigma b4})$.

The RH contract has the same TT constraints and LL constraints and the PC of principal is

$$\alpha \sum_\sigma E_{s|\sigma}PC_\sigma + (1-\alpha)E_{s|N}PC_N$$

The maximum payoff on RH branch is

$$EU_{after}(\alpha, R_{after}) = \alpha \sum_\sigma \pi_\sigma E_{s|\sigma}U_\sigma(R_{after}) + (1-\alpha)EU_N(R_{after}).$$

In more detail

$$\begin{aligned}
EU_{b4}(R_{b4}) &= \pi_G E_{s|G} U_G + \pi_B E_{s|B} U_B \\
&= \pi_G \max_{R_G} [E_{s|G} U_G | E_{s|G} PC_G, TT_G, LL_G] + \pi_B \max_{R_B} [E_{s|B} U_B | E_{s|B} PC_B, TT_B, LL_B] \\
&= \pi_G E_{s|G} U_G (R_{Gb4}) + \pi_B E_{s|B} U_B (R_{Bb4}) \\
&\leq \max_{R_G, R_B} [(\pi_G E_{s|G} U_G + \pi_B E_{s|B} U_B) \mid (\pi_G E_{s|G} PC_G + \pi_B E_{s|B} PC_B), TT_G, LL_G, TT_B, LL_B] \\
&\equiv EU_{after}(1, R_G(1), R_B(1)) \\
&\leq \max_{\alpha} [EU_{after}(\alpha, R_G(\alpha), R_B(\alpha)) | 0 \leq \alpha \leq 1]
\end{aligned}$$

where $R_\sigma(\alpha)$ is the optimal RH branch repayment at the fixed value of α. In fact the first inequality is strict since for $\alpha = 1$ there is cross-subsidisation in the participation constraint with the principal making a loss in the bad signal state and a gain in the good signal state. $\alpha = 1$ is optimal if $EPC_\sigma = 0$. In other cases ($EPC_\sigma \lesseqgtr 0$) optimally $\alpha < 1$ in which case the second inequality is then also strict. Incidentally in these cases there is also cross subsidisation in the different parts of the ex-ante participation constraint but now between the signal and no signal states. ■

E More than two income levels

The contract problem ($\mathcal{P}_{\mathcal{EC}_3}$) is to minimise the enforcement costs (audit cost after a low or intermediate report and the signal acquisition cost (40)) subject to the participation constraint (41) and the incentive constraints (37), (38) and (39).

Proof of Proposition 9. The proof proceeds in steps. We first show that the incentive constraint (37) binds. Then that $m_M = m_{M\sigma} = 0$. Further that also the incentive constraint (39) binds. Last we prove that (38) is slack. This allows us to set the contract problem as one in two states, where the top two, H and M, are pooled into a single state. With suitable modifications, the remaining proofs (Proof of Proposition 10, of Lemma 2 and of Proposition 11) closely mirror those for the two-state case.

1. *The set of incentive constraints (37) binds.*
   It follows from Border and Sobel (1987) that the bottom non-monitored state has a binding incentive constraint with the top monitored state. So if $m_M > 0$, (37) must bind. Conversely, if $m_M = 0$, repayments must be pooled in the top two states and $R_{\hat{M}.} = R_{\hat{H}.}$.

2. $m_M = m_{M\sigma} = 0$ (this is proved below for the no signal case, but the proof is analogous for the with signal case).
   The rationale is that when $m_M > 0$ and $R_{\hat{M}.} < f_M$ there are savings to be made on audit cost by reducing m_M and raising $R_{\hat{M}.}$ in a way which has a neutral effect on the participation constraint (keeping $R_{\hat{M}.}(1 - m_M)$ constant), but slackens (39), reducing the enforcement cost. This goes on until we hit one of the two corners, i.e., either $m_M = 0$, or $R_{\hat{M}.} = f_M$ or both.

Suppose we first reach $m_M = 0$ and $R_{\hat{M}\cdot} < f_M$. Then $R_{\hat{M}M}$ is irrelevant and from (37) binding $R_{\hat{M}\cdot} = R_{\hat{H}\cdot}$. If in constraint (39) we replace $R_{\hat{H}\cdot}$ with $R_{\hat{M}\cdot}$ and rearrange, this becomes $R_{\hat{M}\cdot} \leq (1 - m_L) f_L + m_L f_M$. If we hit $R_{\hat{M}\cdot} = f_M$ with $m_M > 0$, we can reduce both $R_{\hat{H}\cdot}$ and m_M,preserving the participation constraint. Doing this, we can attain $m_M = 0$. To see this, replace $R_{\hat{H}\cdot}$ by f_M, $R_{\hat{M}\cdot}$ by f_M and m_M by zero in the participation constraint, to get

$$\pi_H f_M + \pi_M f_M + \pi_L f_L - I - (\alpha \Sigma_\sigma \pi_{L\sigma} m_{L\sigma} + (1 - \alpha) \pi_L m_L) c_m - \alpha c_a \geq EP_3.$$

This is positive by assumption and shows that there exists a value of $R_{\hat{H}\cdot} \geq f_M$ with $m_M = 0, R_{\hat{M}\cdot} = f_M$ which satisfies the participation constraint.

3. *The set of incentive constraints (39) binds* (this is proved below for the no signal case, but the proof is analogous for the with signal case).

   Consider constraints (38) and (39). Suppose we have an optimum with (39) slack and (as we know from Border and Sobel) $m_L > 0$, since the lowest state always has positive monitoring. The optimum must satisfy both the above inequalities and the participation constraint.

   Keep all the $R_{\hat{H}\cdot}, R_{\hat{M}\cdot}$ fixed, so the right hand sides of the incentive constraints are all fixed. Reducing m_L will keep the second inequality valid for small changes but could violate the first inequality (if it was binding initially). Suppose it was binding initially. Then it is possible to reduce $R_{\hat{H}\cdot}$ too to keep (38) binding. This will also slacken (37). The effect, if possible, is to reduce agency cost. Reducing $R_{\hat{H}\cdot}$ and m_M also has implications for the participation constraint.

   Can you keep doing this until (39) binds at the given $R_{\hat{M}\cdot}$? If so to keep (38) binding there will be finite changes $\Delta R_{\hat{H}\cdot\sigma}, \Delta m_{L\sigma}$ such that

$$(f_H - f_L)(1 - m_L - \Delta m_L) = f_H - R_{\hat{H}\cdot} - \Delta R_H, \tag{50}$$

$$(f_M - f_L)(1 - m_L - \Delta m_L) = f_M - R_{\hat{M}\cdot}. \tag{51}$$

   (50) will certainly hold if $(f_H - f_L) \Delta m_L \geq \Delta R_{\hat{H}\cdot}$ or $\Delta R_{\hat{H}\cdot} / \Delta m_L \leq f_H - f_L$. Within a given signal situation the participation constraint is $\pi_H R_{\hat{H}\cdot} + \pi_M R_{\hat{M}\cdot} + \pi_L (f_L - m_L c_m)$. To prevent this falling with simultaneous reductions in $R_{\hat{H}\cdot}$ and m_L needs $\pi_H \Delta R_{\hat{H}\cdot} \geq \pi_L c_m \Delta m_L$ or $\Delta R_{\hat{H}\cdot} / \Delta m_L \geq \pi_L c_m / \pi_H$. So we need both $\Delta R_{\hat{H}\cdot} / \Delta m_L \leq f_H - f_L$ and $\Delta R_{\hat{H}\cdot} / \Delta m_L \geq \pi_L c_m / \pi_H$. We can find such simultaneous changes $\Delta R_{\hat{H}\cdot}, \Delta m_L$ if $\pi_H (f_H - f_L) \geq \pi_L c_m$. We can reduce m_L and $R_{\hat{H}\cdot}$ to reach a solution where (39) binds if $\pi_H (f_H - f_L) \geq \pi_L c_m$ so long as there are $0 \leq m_L^* \leq 1, R_{\hat{H}\cdot}^* < f_H$ such that

$$(f_H - f_L)(1 - m_L^*) \leq f_H - R_{\hat{H}\cdot}^*$$

$$(f_M - f_L)(1 - m_L^*) = f_M - R_{\hat{M}\cdot}$$

   which would imply

$$\begin{aligned}(f_M - f_L)(1 - m_L^*) &= f_M - R_{\hat{M}\cdot} \\ &\leq f_M - f_L \leq f_H - f_L\end{aligned}$$

But this is always true. So it is always possible to reduce m_L sufficiently to make (39) bind.

4. *The set of incentive constraints (38) is slack.*
Setting $m_M, m_{M\sigma} = 0$, $R_{\hat{H}\cdot} = R_{\hat{M}\cdot}$ and (39) binding, the remaining incentive constraints (38) and (39) become respectively:

$$\begin{aligned} R_{\hat{M}\cdot} - f_L &\leq m_L (f_H - f_L), R_{\hat{M}\cdot\sigma} - f_L \leq m_{L\sigma} (f_H - f_L) \\ R_{\hat{M}\cdot} - f_L &= m_L (f_M - f_L), R_{\hat{M}\cdot\sigma} - f_L = m_{L\sigma} (f_M - f_L) \end{aligned}$$

whence we deduce that if the last set is binding, the first must be slack.

■

Proof of Proposition 10. Suppose that $0 < \alpha < 1$ and $c_a > 0$.

1. An optimum cannot have all $m_{L\sigma} = m_L = 1, \sigma = G, B$. If it did, the participation constraint (43) would be slack ($((\pi_H + \pi_M) f_M + \pi_L f_L - I - \pi_L c_m - \alpha c_a > 0$ by assumption) and a rent would be needlessly left to the principal. Similarly, it cannot have all $m_{L\sigma} = m_L = 0, \sigma = G, B$, since then the expected repayments would not meet the reservation level of the principal ($f_L - I - \alpha c_a < 0$). Thus at least one audit probability must be positive.

2. Suppose that there are $m_{L\sigma}, m_N > 0, \sigma = G, B$, and $m_{LG} < 1$ such that the participation constraint (43) holds. This is not an optimum because it is possible to lower m_{LB} (i.e. lower $R_{\hat{M}\cdot B}$) and increase m_{LG} (i.e. increase $R_{\hat{M}\cdot G}$) reducing the expected cost (42) and slackening the participation constraint (43). Indeed because $\pi_{HG} + \pi_{MG} > \pi_{HB} + \pi_{MB}$ the increase in m_{LG} ($R_{\hat{M}\cdot G}$) and reduction in m_{LB} ($R_{\hat{M}\cdot B}$) allows an increase in the expected revenues after a good signal larger than the fall in revenues after the bad signal. Similarly this shift reduces the with signal expected audit cost since $\pi_{LG} < \pi_{LB}$. This can be seen in the following expression:

$$\begin{aligned} &\frac{1}{f_H - f_L} \{[(\pi_{HG} + \pi_{MG}) (f_M - f_L) - \pi_{LG} c_m] \, dR_{\hat{M}\cdot G} \\ &+ ((\pi_{HB} + \pi_{MB}) (f_M - f_L) - \pi_{LB} c_m) \, dR_{\hat{M}\cdot B}\cdot\} \end{aligned}$$

Since $[(\pi_{HG} + \pi_{MG}) (f_H - f_L) - \pi_{LG} c_m] - [(\pi_{HB} + \pi_{MB}) (f_H - f_L) - \pi_{LB} c_m] > 0$, the participation constraint is slackened by an increase in m_{LG} matched by an equal reduction in m_{LB}. This allows a further reduction in m_{LB} and thus a reduction of the objective function. So an optimum cannot have $m_{LG} < 1$ and $m_{LB} > 0$ and must at least entail one of $m_{LG} = 1$ or $m_{LB} = 0$.

3. Suppose that $m_{LG} = 1$ but $1 > m_{LB} > 0$. Then if $m_L < 1$, it is possible to vary m_L and m_{LB} keeping the participation constraint constant

$$dm_L = -\frac{\alpha [(\pi_{HB} + \pi_{MB}) (f_M - f_L) - \pi_{LB} c_m]}{(1 - \alpha) [(\pi_H + \pi_M) (f_M - f_L) - \pi_L c_m]} dm_{LB} < 0 \tag{52}$$

Thus, m_{LB} and m_L must vary in opposite directions to satisfy the participation constraint. The effect of such variations on the objective function (42) is $dEC = \alpha c_m [-\pi_{LG} dm_N + \pi_{LB}(dm_L - dm_L)] + \pi_L c_m dm_L$, which, using (52), becomes $dEC = \alpha c_m \frac{\rho_3(f_M - f_L)}{(\pi_H+\pi_M)(f_M-f_L)-\pi_L c_m} dm_{LB}$, where $\rho_3 \equiv [\pi_{LB}(\pi_{HG}+\pi_{MG}) - \pi_{LG}(\pi_{HB}+\pi_{MB})]$ is the equivalent for the 3-states case of the correlation index ρ. Since the coefficient of dm_{LB} is positive, it is best to reduce m_B and raise $m_{L.}$ as far as possible.

(a) One possibility is to hit $m_{LB} = 0$ with $m_L > 0$. This proves the result.

(b) Alternatively $m_L = 1$ with $m_{LB} > 0$ still. In this last case the optimisation problem becomes:

$$\begin{aligned} &\min \alpha \left[c_a - \pi_{LB} c_m (1 - m_{LB})\right] + \pi_L c_m \\ \text{s.t. } &- \alpha \left\{(1 - m_{LB}) \left[(\pi_{HB} + \pi_{MB})(f_M - f_L) - \pi_{LB} c_m\right] + c_a\right\} \\ &+ ((\pi_H + \pi_M) f_M + \pi_L f_L - \pi_L c_m) = I \end{aligned}$$

It is then possible to vary α and m_B keeping the participation constraint constant

$$d\alpha = \frac{\alpha \left\{\left[(\pi_{HB} + \pi_{MB})(f_M - f_L) - \pi_{LB} c_m\right]\right\}}{(1 - m_{LB})\left[(\pi_{HB} + \pi_{MB})(f_M - f_L) - \pi_{LB} c_m\right] + c_a} dm_{LB} > 0 \tag{53}$$

The effect on the objective function is $dEC = [c_a - \pi_{LB} c_m (1 - m_{LB})] d\alpha + \alpha \pi_{LB} c_m dm_{LB}$, whence, using (53) gives

$$dEC = \frac{\alpha c_a (\pi_{HB} + \pi_{MB})(f_M - f_L)}{(1 - m_{LB})\left[(\pi_{HB} + \pi_{MB})(f_M - f_L) - \pi_{LB} c_m\right] + c_a} dm_{LB} > 0.$$

It is possible to reduce m_{LB} to zero whilst keeping $\alpha > 0$ since, with $m_{LG} = m_L = 1$, $m_{LB} = 0$, the participation constraint becomes $\alpha EPC_\sigma + (1 - \alpha)((\pi_H + \pi_M) f_M + \pi_L f_L - I - \pi_L c_m) \geq 0$. Since $((\pi_H + \pi_M) f_M + \pi_L f_L - I - \pi_L c_m) > 0$, this can be satisfied with $\alpha > 0$.

4. Alternatively suppose the optimum has $m_{LB} = 0$ but $0 < m_{LG} < 1$ and $m_L > 0$. It is then possible to reduce m_L, raise m_{LG} so as to keep the participation constraint constant: $dm_L = -\frac{\alpha\{(\pi_{HG}+\pi_{HB})(f_M-f_L)-\pi_{LG}c_m\}}{(1-\alpha)\{(\pi_H+\pi_M)(f_M-f_L)-\pi_L c_m\}} dm_{LG}$. The change in objective (42) is $dEC = \alpha c_m \frac{-\rho_3(f_H - f_L)}{(\pi_H+\pi_M)(f_M-f_L)-\pi_L c_m} dm_{LG} < 0$. So it is best to raise m_{LG} to maximum and reduce m_L.

(a) One possibility is that we hit $m_{LG} = 1$ with $m_L > 0$, which proves the result.

(b) The other possible case has $m_L = 0$ with $m_{LG} < 1$ still. In this last case the optimisation problem becomes

$$\begin{aligned} &\min \alpha (c_a + \pi_{LG} m_{LG} c_m) \\ \text{s.t. } &\alpha \left\{\left[(\pi_{HG} + \pi_{HB})(f_M - f_L) - \pi_{LG} c_m\right] m_{LG} - c_a\right\} = I - f_L. \end{aligned}$$

Varying m_{LG} and α so as to keep the participation constraint constant gives $dm_{LG} = -\frac{((\pi_{HG}+\pi_{HB})(f_M-f_L)-\pi_{LG}c_m)m_{LG}-c_a}{\alpha[(\pi_{HG}+\pi_{HB})(f_M-f_L)-\pi_{LG}c_m]}d\alpha$. Using the resulting dm_{LG} in the objective function $dEC = (c_a + c_m \pi_{LG} m_{LG})\, d\alpha + \alpha c_m \pi_{LG} dm_{LG}$ gives $dEC = \frac{(\pi_{HG}+\pi_{HB})(f_M-f_L)c_a}{(\pi_{HG}+\pi_{HB})(f_M-f_L)-\pi_{LG}c_m}d\alpha > 0$. So it is best to reduce α and raise $m_{LG.}$ as far as possible. In fact we can attain $m_{LG} = 1$ with $\alpha > 0$.

Hence the optimum must involve $m_{LG} = 1, m_{LB} = 0$. ■

Proposition 10 is valid if $\alpha > 0$. If $\alpha = 0, m_{L\sigma}$ have no impact on participation constraint or objective and so without loss of generality we can also set them to these values if $\alpha = 0$.

To prove that the results in Proposition 3 (3) extend to more than two states, we use the following lemma.

Lemma 2 *If $0 \leq \alpha < 1$ and $0 < m_L < 1$ are optimal, if $EPC_\sigma \neq 0$, a simultaneous variation in α and m_L keeping the participation constraint constant changes expected enforcement cost according to*

$$dEC = \frac{\{(\pi_H + \pi_M)\, c_a - [\pi_{LB}\,(\pi_{HG} + \pi_{MG}) - \pi_{LG}\,(\pi_{HB} + \pi_{MB})]\, c_m\}\,(f_M - f_L)}{(\pi_H + \pi_M)\,(f_M - f_L) - \pi_L c_m} d\alpha,$$

whose sign depends on the sign of $(\pi_H + \pi_M)\, c_a - \rho_3 c_m \equiv NC_{a3}$.

Proof. Consider Programme $\mathcal{P}'_{\mathcal{EC}}$. The participation constraint is

$$\alpha\,\{(\pi_{HG} + \pi_{MG})\, f_M + (1 - \pi_{HG} - \pi_{MG})\, f_L - I - \pi_{LG} c_m - c_a\} +$$
$$+ (1 - \alpha)\,\big\{(\pi_H + \pi_M)\, R_{\hat{M}.} + \pi_L\,(f_L - m_L c_m) - I\big\} = 0$$

which can also be written as:

$$\alpha\,\{(\pi_{HG} + \pi_{MG})\, f_M + (1 - \pi_{HG} - \pi_{MG})\, f_L - I - \pi_{LG} c_m - c_a\} +$$
$$+ (1 - \alpha)\,\{(\pi_H + \pi_M)\,(f_M - f_L)\, m_L + (\pi_H + \pi_M)\, f_L + \pi_L\,(f_L - m_L c_m) - I\} = 0$$

Suppose $0 \leq \alpha < 1$ and $0 < m_L < 1$ are optimal. When $EPC_{\sigma_3} \neq 0$ simultaneous variation in α, m_L to keep the participation constraint constant must satisfy

$$\{EPC_{\sigma_3} - (PC_N - I)\}\, d\alpha + (1 - \alpha)\,(p_{HM}\,(f_M - f_L) - \pi_L c_m)\, dm_L = 0$$

$$\frac{dm_L}{d\alpha} = -\frac{EPC_{\sigma_3} - (PC_N - I)}{(1 - \alpha)\,[(\pi_H + \pi_M)\,(f_M - f_L) - \pi_L c_m]} \tag{54}$$

While the sign of the denominator is positive by Assumption 36, the sign of the numerator is ambiguous. However, it can be inferred from the sign of EPC_{σ_3}. For $\alpha \in (0, 1)$, to satisfy the participation constraint with equality, $PC_N - I \lesseqgtr 0$ if $EPC_{\sigma_3} \gtreqless 0$. Thus, if:

- $EPC_{\sigma_3} > 0$, $\frac{dm_L}{d\alpha}|_{PC} < 0$;
- $EPC_{\sigma_3} < 0$, $\frac{dm_L}{d\alpha}|_{PC} > 0$.

The simultaneous variation in m_L, α changes the expected enforcement cost by $(c_a + \pi_{LG} c_m - \pi_L m_L c_m)\, d\alpha + (1-\alpha)\, \pi_L c_m dm_L$, i.e., using 54,

$$dEC = \frac{\{(\pi_H + \pi_M)\, c_a - [\pi_{LB}\,(\pi_{HG} + \pi_{MG}) - \pi_{LG}\,(\pi_{HB} + \pi_{MB})]\, c_m\}\,(f_M - f_L)}{(\pi_H + \pi_M)\,(f_M - f_L) - \pi_L c_m} d\alpha. \tag{55}$$

- If $EPC_{\sigma_3} = 0$, $PC = (1-\alpha)(m_L\,(\pi_H\,(f_H - f_L) - \pi_L c_m) - I + f_L)$ and given $\alpha < 1$ we must have $m_L\,((\pi_H + \pi_M)\,(f_M - f_L) - \pi_L c_m) - I + f_L = 0$. In this case no marginal compensation between m_L, α is possible.

■

Proof of Proposition 11. If $NC_{a_3} < 0$, from (55) $\frac{dEC}{d\alpha} < 0$: enforcement costs are minimised by raising α. However, the change in m_L following a change in α necessary to preserve the participation constraint depends on the sign of EPC_{σ_3}. Three cases can arise:

1. If $EPC_{\sigma_3} > 0$, from (54) $\frac{dm_L}{d\alpha}|_{PC} < 0$. Thus m_L and α vary in opposite directions to satisfy the participation constraint. To minimise the enforcement cost, it is best to reduce m_L as far as possible resulting in $m_L = 0$ ($R_{\hat{M}.} = f_L$) and raise α just far enough to satisfy the participation constraint ($\alpha = \underline{\alpha}_3$). Setting $\alpha = 1$ (and m_L interior) is not optimal as it would give a slack participation constraint. The enforcement costs are $\underline{\alpha}_3\,(c_a + \pi_{LG} c_m)$.

2. If $EPC_{\sigma_3} = 0$, since $NC_{a_3} < 0$ now it is best to set $\alpha = 1$ and then m_L is immaterial. Enforcement costs are $c_a + \pi_{LG} c_m$.

3. If $EPC_{\sigma_3} < 0$, $\frac{dm_L}{d\alpha}|_{PC} > 0$. Thus m_L and α vary in the same direction to satisfy the participation constraint and α has to be increased to minimise the enforcement cost. Setting $\alpha = 1$ is not optimal as it would violate participation constraint. Thus m_L must be raised as far as possible, giving $m_L = 1$ and α must be raised sufficiently to satisfy the participation constraint ($\alpha = \bar{\alpha}_3$). The enforcement costs are $\bar{\alpha}_3\,(c_a + \pi_{LG} c_m) + (1 - \bar{\alpha}_3)\,\pi_L c_m$.

■

F Acknowledgements

We are grateful to two anonymous referees, Sandeep Baliga, Alberto Bennardo, Gabriella Chiesa, Marcello D'Amato, Paolo Garella, Tracy Lewis, Dilip Mookherjee, Marco Pagano, Nicola Persico, Salvatore Piccolo, Ailsa Roell, Maria Grazia Romano, Joel Sobel, Francesca Toscano, Zaifu Yang, the participants to the CSEF-IGIER Symposium on Economics and Institutions in Capri, the ASSET Conference in Padua, the RES Conference in Surrey, the 50th SIE Conference in Rome, the EARIE Conference in Toulouse, as well as the seminar participants at the universities of Bologna and Naples. The usual disclaimer applies.

References

[1] Andreoni, J., Erard, B. and J. Feinstein (1998), "Tax Compliance," *Journal of Economic Literature*, **36**, 818-860.

[2] Border, K.C. and J. Sobel (1987), "Samurai Accounting: A Theory of Auditing and Plunder," *Review of Economic Studies*, **54**, 525-540.

[3] Cella, M. (2008), "Informed Principal with Correlation," *Games and Economic Behavior,* **64**, 433-456.

[4] Chander, P. and L.L. Wilde (1998), "A General Characterization of Optimal Income Tax Enforcement," *The Review of Economic Studies*, **65**, 165-183.

[5] Chew, M. C.(1973), "On Pairing Observations from a Distribution with Monotone Likelihood Ratio," *Annals of Statistics*, **1** (3) , 433-445

[6] Crémer, J. and F. Khalil (1992), "Gathering Information before Signing a Contract," *American Economic Review*, **82** , 566-578.

[7] Dewatripont, M. (1988), "Commitment through Renegotiation-Proof Contracts with Third Parties," *Review of Economic Studies*, **55** (3), 377-89.

[8] Erard, B. and J. Feinstein (2010), "Econometric Models for Multi-Stage Audit Processes: An Application to the IRS National Research Program," in *Developing Alternative Frameworks for Explaining Tax Compliance*, edited by James Alm, Jorge Martinez-Vazquez and Benno Torgler, Routledge, 113-37.

[9] Gale, D. and M. Hellwig (1985), "Incentive-compatible debt contracts: the one period problem," *Review of Economic Studies*, **52**, 647-663.

[10] Hart, O. (1995), "*Firms, Contracts and Financial Structure*," Clarendon Press, Oxford.

[11] Khalil, F. (1997), "Auditing without Commitment," *RAND Journal of Economics*, **28**, 629-640.

[12] Khalil, F. and B. Parigi (1998), "Loan Size as a Commitment Device," *International Economic Review*, **39**, 135-150.

[13] Krasa, S. and A.P. Villamil (2000), "Optimal Contracts When Enforcement is a Decision Variable," *Econometrica*, **68**, 119-134.

[14] Lewis, T.R. and D.E.M. Sappington (1997), "Information Management in Incentive Problems," *Journal of Political Economy,* **105**, 796 821.

[15] Macho-Stadler, I. and D. Pérez-Castrillo (2002), "Auditing with Signals," *Economica*, **69**, 1–20.

[16] Maskin, E. and J. Tirole (1990), "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, **58**, 379-409.

[17] Maskin, E. and J. Tirole (1992) "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, **60**, 1-42.

[18] Menichini, A. and P. Simmons (2012), "Auditing, Insurance and Bargaining Power: The Case with Two States of Nature," mimeo.

[19] Mookherjee, D. and I. Png (1989), "Optimal Auditing, Insurance and Redistribution," *Quarterly Journal of Economics*, **104**, 399-415.

[20] Reinganum, J. F. and L.L. Wilde (1985), Income Tax Compliance in a Principal Agent Framework," *Journal of Public Economics*, **26**, l-18.

[21] Reinganum, J. F. and L.L. Wilde (1986), "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance," *International Economic Review*, **27**, 739-760.

[22] Scotchmer, S. (1987), "Audit Classes and Tax Enforcement Policy," *American Economic Review*, **77**, 229-33.

[23] Thuronyi, V. (1996), "Presumptive Taxation," in V. Thuronyi (ed.), *Tax Law Design and Drafting*, vol. 1, Washington D.C., International Monetary Fund.